

## **THE MARKING OF ENGLISH ESSAYS**

# INTERNATIONAL INSTITUTE EXAMINATIONS ENQUIRY

## Members of the Committee

SIR MICHAEL SADLER, K.C.S.I., C.B., LL.D. (*Chairman*), Sometime Master of University College, Oxford, and Vice-Chancellor of the University of Leeds.

P. B. BALLARD, M.A., D.Lit., Sometime Inspector in the Education Department of the London County Council.

C. DELISLE BURNS, M.A., D.Lit., Sometime Stevenson Lecturer in Citizenship in the University of Glasgow.

CYRIL BURT, M.A., D.Sc., Professor of Psychology in the University of London.

F. CLARKE, M.A., D.Litt., Professor of Education in the University of London.

H. R. HAMLEY, M.A., M.Sc., Ph.D., Professor of Education in the University of London.

SIR PHILIP HARTOG, K.B.E., C.I.E., LL.D. (*Director*), Sometime Vice-Chancellor of the University of Dacca and Chairman of the Auxiliary Committee on Education of the Indian Statutory Commission.

SIR PRINCE NESS, D.Sc., Litt.D., Emeritus Professor of Education in the University of London.

C. SPEARMAN, LL.D., F.R.S., Emeritus Professor of Psychology in the University of London.

MISS ETHEL STANDWICK, O.B.E., M.A., High Mistress of St. Paul's Girls' School, London.

GODFREY H. THOMSON, D.Sc., Professor of Education in the University of Edinburgh.

---

## Other publications of the International Institute Examinations Enquiry Committee

AN ENGLISH BIBLIOGRAPHY OF EXAMINATIONS (1900-1932), by Mary C. Champneys, with a Foreword by Sir Michael Sadler and Sir Philip Hartog (pp. xxiv, 141), 1934. Price 5s.

AN EXAMINATION OF EXAMINATIONS, by Sir Philip Hartog and Dr. E. C. Rhodes (Second Edition, Fourth Impression, pp. 81), 1936. Price 1s. (First Published 1935.)

ESSAYS ON EXAMINATIONS, by Sir Michael Sadler, A. Abbott, P. B. Ballard, Cyril Burt, C. Delisle Burns, Sir Philip Hartog, C. Spearman, and S. D. Stirk (pp. xii, 168), 1936. Price 5s.

THE MARKS OF EXAMINERS, by Sir Philip Hartog and Dr. E. C. Rhodes, with a Memorandum by Professor Cyril Burt (pp. xix, 344), 1936. Price 8s. 6d.

A CONSPECTUS OF EXAMINATIONS IN GREAT BRITAIN AND NORTHERN IRELAND, by Sir Philip Hartog with the assistance of Gladys Roberts (pp. xiv, 182), 1937. Price 3s. 6d.

INTERNATIONAL INSTITUTE EXAMINATIONS ENQUIRY

---

THE  
**MARKING OF ENGLISH  
ESSAYS**

A Report on an Investigation carried out by a  
Sub-Committee of the International Institute  
Examinations Enquiry Committee consisting of  
Sir PHILIP HARTOG (*Chairman*), Dr. P. B. BALLARD,  
Dr. P. GURREY, Professor H. R. HAMLEY, and  
Dr. C. EBBLEWHITE SMITH

WITH A STATISTICAL REPORT BY  
DR. C. EBBLEWHITE SMITH

MACMILLAN AND CO., LIMITED  
ST. MARTIN'S STREET, LONDON

1941

MADE AND PRINTED IN GREAT BRITAIN  
AT THE CHAPEL RIVER PRESS  
ANDOVER, HANTS

## PREFACE

The object of this Report, which we present to the public on behalf of our Committee, is constructive rather than critical. The last chapter (Chapter VIII) contains a number of Recommendations for future action to which we draw the attention of our readers. Those Recommendations are based in part on previous work of the members of the Sub-Committee by whom the Report has been drawn up, in part on the results of the investigation which it records, and on the inability of examiners accustomed to the present methods of teaching English and examining in English composition to attain a satisfactory degree of consistency in the marking of English essays. But we must leave the book and its Recommendations to speak for themselves. They deal with fundamentals in the teaching of the mother-tongue—the most important element in the national education, as it was until recent years the most neglected—and in the examinations in that subject.

The experimental part of the investigations which we record was begun in 1935 and completed in 1938. But the issue of this book has been delayed owing to the illness of some of the persons concerned and the exigencies of the war. Of the five persons who formed the Sub-Committee which supervised the investigation only Dr. Ballard and the Director were available for the drafting of Chapters VII and VIII, and of the Notes at the end of the book.<sup>1</sup> We have appended below the acknowledgments of our Committee to the numerous bodies and persons who have assisted directly or indirectly in the present investigation; but we wish to record, in this sixth and last book to be published by our Committee, our primary obligations to those bodies and persons who made our undertaking possible in the first instance. Those bodies were the Carnegie Corporation, the Carnegie Foundation, and the International Institute of Teachers College, Columbia University, New York. The movement was initiated at an international conference, held at Eastbourne, under the auspices of these three American

<sup>1</sup> Dr. Gurrey rejoined the Royal Air Force early in the war; Dr. Eblewhite Smith left England to fill a Chair of Education at Winnipeg at the request of the Government of Manitoba; and Professor Hamley has gone recently to Irak on a mission for the British Council.

organisations, in May, 1931; and it was followed by similar conferences at Folkestone in June, 1935, and at Dinard, in France, in September, 1938. All the conferences were organised by Dr. Paul Monroe, Director of the International Institute, and held under his able chairmanship. Dr. Monroe also edited their published proceedings.<sup>1</sup> The conferences were attended not only by representatives of the several European committees,<sup>2</sup> but on each occasion also by a contingent of outstanding representatives of the United States. The proceedings contain verbatim reports of the speeches and discussions, many of them lively, and some of them brilliant. They give a first-hand impression of features in the examination systems of different countries not to be found elsewhere. The meetings have left in the minds of those of us who attended them unforgettable memories, not alone of the discussions, but of the hospitality and understanding of our American hosts; and above all of Dr. Monroe, that veteran figure in American education, who has been the life and soul of the movement, and of Dr. F. P. Keppel, the statesmanlike President of the Carnegie Corporation, which supplied the funds for the scheme through the International Institute. Dr. Monroe related at Dinard how the movement originated, some 10 or 11 years ago, in a conversation between Dr. Keppel and himself, though the parts played by the two men have been fused in their common recollections.<sup>3</sup>

It was by the desire of the French representatives that the last conference was held in France. It was held, under the shadow of war, in the tense fortnight preceding the days of Munich; but the spirit of the conference was one of international co-operation. Of its memories none will be more permanent than that of the hours spent on a lovely bay in Brittany, in the country-house of M. Célestin Bouglé, Director of the Ecole Normale Supérieure in Paris, whose recent death we deplore. At the end of the conference, on the suggestion of Professor Drever, of Edinburgh, an informal international committee was set up

<sup>1</sup> The three volumes of Proceedings were published in 1931, 1936, and 1939, respectively, by the Bureau of Publications, Teachers College, Columbia University, New York. The short title of each volume is "Conference on Examinations," followed by the name of the place in which it was held.

<sup>2</sup> Committees parallel to our own were established in the following countries: Finland, France, Germany, Norway, Scotland, Sweden and Switzerland. These committees have simultaneously investigated such problems of examinations as they regarded as urgent and have published books or memoirs on the subject. The German committee ceased to function not long after the establishment of the Nazi régime.

<sup>3</sup> See the Proceedings of the Dinard Conference, pp. 8-11.

(with Professor Drever as convener) to arrange for a further meeting at a date and place to be fixed in the future.

There is one more thing to report. The Carnegie Corporation spontaneously offered the English Committee a sum up to the limit of \$10,000, to be matched by a like amount from England, as the nucleus of a fund for educational research in this country, not limited to questions relative to examinations. With the approval of the American donors it was decided that the fund should be devoted to grants for approved researches, open to anyone in England and Wales (on the model of those made by the Royal Society) and to be administered by the University of London Institute of Education with the assistance of an Advisory Council. Before the war, considerable progress had been made with the scheme. Negotiations were being carried on with a view to enlist the support of the local education authorities of England and Wales, through the important associations of officials and others connected with them ; of the great associations of teachers such as the National Union of Teachers and the four associations of Secondary Teachers ; and of other bodies such as the National Institute of Industrial Psychology. With the onset of war, the possibility of setting up a complex system of administration and finance of the kind contemplated disappeared ; but the Leverhulme Trustees very generously allowed their offer of £1,000 for the scheme to stand, and the American supporters immediately cabled a like sum to match this gift. These sums have now been transferred by our Committee to the Institute of Education, which is, for the moment, located at University College, Nottingham. It is understood that the Institute, of which Professor Fred Clarke is the Director, will now be prepared to make grants from the capital for educational researches of which the utility at the present time is recognised ; and it is hoped that further donations for the same purpose will be received. The fund is to be called the Carnegie Fund for Educational Research.

The joint action of Dr. Monroe, of Dr. Keppel, and of Dr. Walter Jessup, President of the Carnegie Foundation, will, we trust, long bear fruit in this sister country to the United States, to which we are bound more than ever by ties of kinship, understanding, affection, and a common tradition of humanity.

---

#### *Acknowledgments*

We have a number of acknowledgments to make on behalf of the main Committee to those who helped our Sub-Committee

in the production of this Report. First and foremost to Dr. Ebblewhite Smith and to his wife, who gave their time without stint or reward to the laborious analysis of statistical results recorded mainly in Chapter III. Their work and that of their colleagues has served as a corner-stone of the conclusions arrived at. To Dr. Ballard, Dr. Percival Gurrey and Professor Hamley, who served with Dr. Smith as members of the Sub-Committee, and to Professors Cyril Burt, Spearman and Godfrey Thomson, who made valuable suggestions at a later stage, we owe sincere gratitude. We desire to acknowledge also the assistance of the University of London Institute of Education, for providing accommodation for the meetings of the examiners; and our obligations to Miss Gladys Roberts, Mr. E. C. Rubidge, Miss Ann Wright and Miss G. Rohatyn, for skilled clerical assistance at various stages of the work, and to Mr. G. B. N. Hartog and Mr. Roger Hartog for the time they voluntarily devoted in the early stages to dealing with the thousands of scripts in our hands, thus avoiding the necessity of engaging additional staff. The Sub-Committee, in the text of the Report, have recorded the great debt which they owe, not only to the educational authorities concerned, but also to the headmasters and headmistresses and English teachers of the schools participating in the enquiry for having carried out the directions for the conduct of the examinations and tests, in faithful accordance with the schemes laid down both for the preliminary and for the final investigations. Without their invaluable co-operation the investigations could not have been carried out. We are also indebted to the National Institute of Industrial Psychology for allowing our Sub-Committee the use of their Group-Intelligence Test No. 133 with its 193 items and for scoring the papers of the candidates.

Finally, we have to express to Mr. P. H. B. Lyon, Headmaster of Rugby, our thanks for communicating to us the results of an interesting test in the marking of school essays carried out by two members of his staff; and to Dr. F. A. Hedcock, Inspector under the London County Council, for a promising investigation on a method of marking French compositions, which however it was impossible to advance to a stage suitable for the publication of the results.

For the Committee,

M. E. SADLER, *Chairman.*

P. J. HARTOG, *Director.*

*April, 1941.*

## TABLE OF CONTENTS

	PAGE
PREFACE BY SIR MICHAEL SADLER AND SIR PHILIP HARTOG . . . . .	v
LIST OF TABLES . . . . .	xiii
CHAPTER I.—INTRODUCTORY . . . . .	1
The main object of the investigation. Importance of the ideas of "audience" and "object" in writing in real life. Introduction into marking of the element "sense" with a defined connotations. (§§ 1-6).	
CHAPTER II.—PRELIMINARY INVESTIGATION . . . . .	3
Outline scheme of this investigation (§ 7). Description of the tests used (§§ 8-12). Conclusions (§§ 13-15).	
CHAPTER III.—INVESTIGATION ON THE MARKING OF ENGLISH ESSAYS, APPROXIMATELY OF SCHOOL CERTIFICATE STANDARD, AND THE RELATION OF THE MARKS TO THOSE AWARDED FOR CERTAIN OBJECTIVE TESTS <sup>1</sup> . . . . .	8
Nature of the school population and description of the tests applied (§§ 16-22). Average age of the pupils (§ 17). Subjects of the six essay-tests directed and undirected and time allowed (§ 18). Length of essays (§ 19). Method of the selection of the material for investigation (§ 20). Mechanical reproduction of the scripts (§ 21). Nature of objective tests (§ 22). Procedure adopted for the marking of the essay-scripts (§§ 23-34). Personnel of the Board of Examiners and of the Sub-Committee of the main committee by whom the marking of the scripts was devised (§ 23). Trial-scripts (§ 27). Categories marked and scale of marking (§ 28). Median mark of scale to denote competency (§ 29). Marking of trial-scripts (§§ 30, 31 and 33). Marking for Literacy (§§ 32 and 34). Marking-sheets (§ 34). Opinions and Recommendations of Examiners (§§ 35-35B). Re-marking of 120 scripts (§ 36).	

<sup>1</sup> The historical treatment of the proceedings of the sub-committee referred to and the Board of Examiners has made it impossible to avoid some overlapping of the sections dealing with different subjects.

**CHAPTER IV.—STATISTICAL REPORT BY DR. C.  
EBBLEWHITE SMITH<sup>1</sup>**

20

Summary of scheme. The main task, the analysis of 34,100 marks, yielded by the first marking of 600 scripts, and 5 objective tests, and 6720 marks yielded by the second marking of 120 scripts (exclusive of the marks for Literacy) (§§ 37–45).

Measure of agreement between examiners; three aspects of this question (§ 46). Differences in "standard" and in "spread" of marks. Use of "standard scores." Use of "cumulative frequency curves" based on standard scores. 1200 such curves drawn, of which 4 per cent. were discontinuous. Sample figures of such curves (§§ 47–58).

Analysis of Variance (§§ 59–84). Conclusion (§ 85).

Factor Analysis (§§ 86–94). Summary of results (§ 95).

Re-marking of scripts (§§ 96–103). Re-marking for Literacy (§ 104). Conclusion (§ 105).

**CHAPTER V.—LITERACY AS A CRITERION AT THE  
SCHOOL CERTIFICATE STAGE**

80

No definition of Literacy suitable for this purpose. The value of the term tested by experiment. Adoption of a three-point scale (§§ 108–9). Disagreements between examiners (§§ 110–115, and Tables XX to XXVIII). Conclusion: The term Literacy has no meaning sufficiently well-recognised for it to be effectively used in judging the scripts of the candidates (§ 116).

Re-marking for Literacy (§§ 117–120). Typical examples of essays regarded by examiners as illiterate or of doubtful literacy (§§ 123–132).

**CHAPTER VI.—EXAMPLES OF ESSAYS SELECTED  
ON THE BASIS OF THE MARKS ALLOTTED TO  
THEM BY THE EXAMINERS**

93

**A. EXAMPLES OF THE BEST ESSAYS.**

Introductory (§ 133). Essay-subject A (§ 134). Essay-subject A1 (§ 135). Essay-subject B (§ 136). Essay-subject B1 (§ 137). Essay-subject C (§ 138). Essay-subject C1 (§ 139).

**B. EXAMPLES OF ESSAYS REGARDED AS JUST "COMPETENT."**

Two essays on subject C1 (§ 140).

**C. EXAMPLES OF ESSAYS IN REGARD TO WHICH THE EXAMINERS  
DIFFER WIDELY IN OPINION.**

Introductory (§ 142). Essay-subject A (§§ 143–4). Essay-subject A1 (§§ 145–6). Essay-subject B (§§ 147–8). Essay-subject B1 (§§ 149–50). Essay-subject C (§§ 151–2). Essay-subject C1 (§§ 153–4).

<sup>1</sup> The general plan of the statistical investigation may be gathered from the List of Tables, pp. xiii–xv, of which the contents are not repeated in this table of contents.

PAGE  
128

## CHAPTER VII.—CONCLUSIONS . . . . .

Though the investigation has yielded results of importance, the statistical report shows that the introduction of the element of Sense into the marks has failed in its purpose of increasing the consistency of the marking (§ 154). The first method used by the statisticians, the drawing and analysis of 1,200 "cumulative frequency curves" based on "standard scores" (§ 155). The second method, the Analysis of Variance. The attempt to distinguish between three components in each mark; the uncertainties of the results (§ 156). The method of Factor Analysis. The "factor-pattern." The "halo-effect." Suggestions of Dr. Burt, Dr. Spearman and Dr. Godfrey Thomson. The examiners have been unable to free themselves from the traditions acquired in teaching and previous examining (§ 157). Interpretation by the Sub-Committee of the failure to secure increased consistency (§ 157A). Use of the terms "absolute merit" and true or standard mark by Edgeworth and Rhodes. Fundamental difficulties to be considered later (§§ 158–9). Complexities of marking examined, starting with the simplest of examples, sums in simple addition. Consistency of marking even in arithmetic only to be secured by introducing conventions on which differences of opinion are legitimate (§ 160). The introduction of heterogeneous elements such as "neatness" into the marking and the irrational consequences (§ 161). Dr. Ballard's analysis of the marking of essays. Methods of marking by categories introduced to distinguish the inconsistencies of marking (§ 162). The categories used at a Special Place Examination criticised (§ 163). A more satisfactory scheme of categories, but still open to criticism (§ 164). The equation of marks for spelling and handwriting with marks for elements like Sense or Sentence Structure an absurdity. Marks for spelling and legibility of handwriting should be kept at all stages distinct from those for the other elements of a written composition (§ 165). The justification for including punctuation with the other categories (§ 166). Categories used in our investigation and order in which they were considered. The category of Sense, marked first. Necessity for its introduction into the class-room. Marks for General Impression allotted last, but not as the sum of the marks for the other categories (§ 167). Modified set of categories recommended by the examiners for future use. The Sub-Committee differ from the recommendations on two points (§ 168). The Sub-Committee regard an 8-point scale as unsuitable for the marking of categories (iii) (iv) and (v) (§ 169 (i)). They also regard it as of great importance to insist on the use of a plan both in teaching and in examining (§ 169 (ii)). Distinction of the categories into those to which universally accepted conventions of correctness apply, and the rest, which include "plan" and "power of expression," covering "felicities of vocabulary." The last phrase launches us on a sea of uncertainties (§ 170). Such uncertainties equally constrain the marking of French compositions by French

these conditions the mean marks of a large number of examiners desiderated by Edgeworth seem to have no real significance (§ 171). Candidates ought neither to be ploughed nor passed by chance. Some of the difficulties of the examining bodies due to their own regulations. A candidate who has been ploughed in English but passed in other subjects should be credited with his passes to enable him to devote more time to his English (§ 172). The essay in the form until recently in universal vogue criticised for its unreliability. Investigation by the Durham University in 1931-32. Criticism of the English standard by the Government Investigators of 1931. New regulations recently introduced with reforms of which the precise effect cannot yet be estimated (§ 173). The separation of tests in English language from tests in Literature. Improvement in questions of detail. Abandonment by some authorities of the term "essay" (§ 174). The subjects set, however, are still often vast or vague, or both. Examples (§ 175). Time allowed for an "essay" or a composition (§ 176). Instructions of one authority (§ 177). Suggested reforms ( §§ 178-9). Postscript on use of the term Literacy. Comments on the single script regarded as "illiterate" by all the examiners (§ 180).	
CHAPTER VIII.—SUMMARY OF RECOMMENDATIONS . . . . .	142
NOTE ON THE ALLOTMENT OF THE COMPETENCY MARK IN THE MARKING FOR GENERAL IMPRESSION, AND ON THE RANGES OF THE MARKS IN THAT CATEGORY . . . . .	145
NOTE ON THE ANALYSIS OF VARIANCE AND FACTOR ANALYSIS . . . . .	149
NOTE ON CERTAIN POINTS IN THE MARKING OF EXAMINERS . . . . .	152
NOTE ON AN INVESTIGATION ON THE MARKING OF ENGLISH COMPOSITION BY DR. B. M. D. CAST	156
APPENDIX I.—OBJECTIVE TESTS, 2 to 5 . . . . .	157
APPENDIX II.—SAMPLE OF MARKING-SHEET FOR ESSAYS . . . . .	164

## LIST OF TABLES

		<small>PAGES</small>
<b>TABLE I</b>	Sample Scores on Objective Tests . . . . .	23
<b>TABLES II (1) TO II (6)</b>	Representative Samples of marks awarded by the different Examiners for Essay-subjects A, A1, B, B1, and C, C1 . . . . .	24-29
<b>TABLE III</b>	Means (for 100 Scripts) of Examiners' Marks . . . . .	30-31
<b>TABLE IV</b>	Standard Deviations (for 100 Scripts) of Examiners' Marks . . . . .	32-33
<b>TABLE V</b>	Sample Standard Scores for Category I (Sense) and Category II (General Impression) for Essay-subject A . . . . .	34
<b>TABLE VI</b>	Skeleton table to illustrate method of calculating variancee . . . . .	44
<b>TABLE VII</b>	Table giving values of $V_1$ , $V_2$ , $V_3$ , $V_1/V_3$ , and $V_2/V_3$ , for the several Essay-subjects and Categories . . . . .	47-48
<b>TABLE VIII</b>	Table giving estimated values of $\sigma_B$ , $\sigma_C$ and $\sigma_E$ for the several Essay-subjects and Categories . . . . .	49
<b>TABLES IX-XIV</b>	Intercorrelations between the marks for each of the Categories for each Essay-subject and between those Marks and the marks for the Objective Tests . . . . .	55-57
<b>TABLE XV</b>	Intercorrelations between the Marks for the Objective Tests . . . . .	58
<b>TABLE XVI</b>	Factor Weightings or Loadings for the different Essay-subjects and Categories and the Objective Tests . . . . .	

TABLES XVII (1) TO XVII (6)	Representative samples of the Second Markings by the different Examiners for Essay-subjects A, A1, B, B1, and C, C1, corresponding to the First Markings in Tables II (1) to II (6). . . . .	61-66
TABLE XVIII	Correlations between First and Second Markings of 120 Scripts . . . . .	67
TABLES XIX (1) TO XIX (6)	Comparison of First and Second Mark- ings for General Impression of Essay- subjects A, A1, B, B1, and C, C1 . . . . .	71-76
TABLE XIX (7)	Number and Magnitude of the Differences between the First and Second Markings of the different Examiners for General Impression . . . . .	77
TABLES XIX (8) AND XIX (9)	Averages of the Differences between the First and Second Markings of the Eight Examiners for General Impression (20 Scripts for each Essay-subject) . . . . .	78
TABLE XX	Numbers of cases in which different numbers of Examiners agree in marking for Literacy . . . . .	82
TABLE XXI	Number of Candidates marked "Yes" for Literacy by all the Eight Examiners	82
TABLE XXII	Percentages of the three awards for Literacy allotted by the different Examiners for each Essay-subject . . . . .	83
TABLE XXIII	Ranges of the percentages for each Examiner of each award for Literacy in each Essay-subject . . . . .	83-84
TABLE XXIV	Percentages of the different awards for Literacy allotted for all the Essay- subjects by the different Examiners . . . . .	84
TABLE XXV	Changes between First and Second Markings for Literacy, classified accord- ing to Essay-subjects and Examiners . . . . .	85-86
TABLE XXVI	Summary of number of changes in Re- marking for Literacy . . . . .	86

	PAGES	
TABLE XXVII	Comparison of percentages of the different awards for Literacy allotted at the First and Second Markings by the different Examiners . . . . .	87
TABLE XXVIII	Table showing the numbers of candidates who received one, two, three, etc., "Noes" in the Marking for Literacy .	87
TABLE XXIX	Percentages of Scripts which received the Mark of Competency (S), of a Higher Mark (S +) and of a Lower Mark (S —) in the First Marking for General Impression (600 scripts) . . . . .	147
TABLE XXX	Ditto, in Re-Marking (120 scripts) .	148

# THE MARKING OF ENGLISH ESSAYS

## CHAPTER I.—INTRODUCTORY

1. It is generally agreed that of all the problems which examiners have to face there is none more difficult than the marking of the English Essay. The main object of the investigations dealt with in this Report was to ascertain whether by the introduction of a new element into the marking the validity and the consistency of the examination could both be improved.

2. For the sake of brevity this element was called "sense." But the precise connotation with which the word was used needs further explanation.

In previous work on the subject by the Director of the Enquiry it was pointed out that in real life a person does not just "write." He writes for a given *audience* and with a given *object* in view, which may be to explain, to persuade, to convey an order, to amuse, to inspire, or indeed to fulfil any other purpose or combination of purposes.<sup>1</sup> The word "sense" in this investigation has been used to express the measure in which the object of a composition, clearly defined in some cases, to be guessed from internal evidence in others, has (in the opinion of the examiners) been attained.

3. The following brief examples were communicated to the examiners in our main investigation to illustrate in what way it was desired that the word "sense" should be utilised in the marking of the scripts.

### EXAMPLE I

#### *Testimonial to a Typist*

Miss X is a lady who has had a training for many years under skilled supervision, and is accomplished in secretarial work of all kinds. I regret, however, to say that her spelling is defective, that she is untrustworthy in her filing, and that, owing to her lack of general education, she cannot write a good letter. I hope that, when she has worked for a year or more under someone who is able to train her properly, she will be of real use. I may add that she has one uncle in Japan, and another in Australia.

<sup>1</sup> See *The Writing of English*, 2nd edn., pp. v and 61, and *Record or Message*. Both these publications are out of print, but it is hoped to reprint them in the same or another form. On this subject, see the book by Dr. P. B. Ballard, *Teaching and Testing English* (Univ. of London Press, 1939).

## EXAMPLE II

*Extract from an Essay on Games*

One of the great advantages of school games is that they give players the capacity to act coolly in a crisis.

## 4. Let us take first Example I.

One experienced and highly qualified examiner in English (not one of the team employed in our actual investigation) was frank enough to say that his first impulse was to mark the nonsense contained in the Testimonial to a Typist with an alpha ; and he said afterwards "I still should have given it a high mark because the person who wrote that could write good sentences in English."

5. The point is crucial. Our examiners and teachers have spent so much time in teaching pupils to write sentences that the majority do not realise that by the age of sixteen it is time to consider whether the combination of the sentences makes "sense." When there is a conflict between "sentences" and "sense" the sentences have it every time. And, if this is the view of teachers and examiners, is it to be wondered at that it also comes to be the view of many otherwise intelligent but docile pupils ? The situation was naïvely and strikingly revealed by an English mistress at a secondary school, when she said, not long ago, to a bright girl of fourteen with a mind of her own, " You are too sensible to write a good English composition."

6. We come to Example II, taken from an actual schoolboy essay—" One of the great advantages of games is that they give players the capacity to act coolly in a crisis." Here the spelling does, we may confess, give a shock, but the sense of the passage is irreproachable.

In the marking of a composition as a whole for the element " Sense " nothing should be deducted for a sentence of this kind. But there is little doubt that with most examiners the faulty spelling would influence not only the marks for spelling but those for " General Impression " ; and some might on the strength of it, even condemn such a writer as " illiterate " —a point with which we shall deal later.

## CHAPTER II.—PRELIMINARY INVESTIGATION

7. A preliminary investigation on the marking of essays by school children of from 13 to 16 was conducted in 1936 and the following outline scheme was adopted.

(i) The main object of the present investigation is to devise, if possible, a method of marking English compositions which shall have both greater "validity" and greater "reliability" than the present methods. A word of explanation with regard to the use of these two terms is necessary. By "validity" is meant the degree of agreement of a measurement with the thing measured; by "reliability" is meant the degree of agreement between any two independent sets of measurements of the same set of things.

(ii) To quote Professor Spearman, "the inter-relations of reliability and validity are one-sided. Low reliability necessarily involves low validity, but the converse is not true. Wherever we find bad agreement between different measurements then we can safely say that the examination is bad. But when the measurements agree we can *not* forthwith say that the examination is good."<sup>1</sup>

The low "reliability" or "consistency" of essay-tests is notorious (see *An Examination of Examinations*, pp. 19-32, and *The Marks of Examiners*, pp. 83-92, 117-141 and 142-147); hence, in accordance with the principle enunciated by Professor Spearman, the validity of the tests is necessarily low.

It seems possible that by adopting a new method of examining, based in part on a new method of teaching, both the validity and the consistency of tests in English composition may be increased.

(iii) Two fundamental questions must be asked in connection with all examinations:—

What am I really trying to test?

Am I testing it in the best way possible?

The examiner, in marking English essays at stages below the School Certificate stage, and up to that stage, has to bear many things in mind at once. He is not testing one element at a time, but more like half a dozen. He has to think, at least, of Spelling, Punctuation, Vocabulary, and Grammar, as well as of the most important element of all—the sense, and in the competition for consideration, the element of sense

<sup>1</sup> The term "consistency" will be used below as a substitute for the term "reliability" as used by Professor Spearman in his note "On the Reliability and Validity of Measurements" in the *Essays on Examinations*, published for the Committee (Macmillan, 1936), from which the passage quoted above is taken.

often comes in easily last. In the present investigation tests will be given of two kinds :—

- (1) Class (A) Tests : tests of spelling, vocabulary, punctuation, grammar, etc.
- (2) Class (B) Tests : tests of composition, that is of the power of the candidate to express himself in English for a clearly defined purpose.

*Details of Class (A) Tests.*—These tests will be of the "new type," i.e. of such a kind as to make the marking mechanical and independent of the personality of the marker.<sup>1</sup>

*Details of Class (B) Tests.*—These will be based on principles defined by Sir Philip Hartog in the *Writing of English and Record or Message*, and in the chapter on *English Composition at the School Certificate Examination ; and the "Write Anything about Something for Anybody" Theory*, included in the *Essays on Examinations* published by the Committee. In every case the candidate will not only be informed of the subject of the composition on which he is required to write, but he will also be required to write with a given audience and a given object in view. He will further be required to make, either before he has written his composition or after doing so, a plan in the form of headings, and consisting of not more than ten or a dozen lines. The investigators will then be asked to consider how far in their judgment each essay is likely to be successful or unsuccessful in attaining the object, taking into consideration the audience for whom it is written. A composition from the first line to the last must be controlled by the purpose of the whole.

Two small illustrations of this principle may be given. Let us suppose, for instance, that a boy has been asked to describe a cricket match for the benefit of a foreign boy who was never seen one. It would be absurd for him to refer to the new l.b.w. rule of the M.C.C. without explanation.<sup>2</sup> On the other hand, if he were writing to describe a match for a member of his own eleven it would be equally absurd for him to explain what the abbreviations l.b.w. and M.C.C. mean.

In many cases a small slip in grammar or punctuation or spelling would not in any way affect the success or failure of the composition. In a test for spelling a boy who wrote "wickett" or "pavillion" should be penalised ; but if he is writing a description of a match to interest a fellow-member of his own cricket-team, the vividness and accuracy of his description and its general effect would not in any way be affected by a mis-spelling of this kind. Nor would it be affected by small slips in grammar. No deduction should be made for a sentence like this :

"Neither of the batsmen *were* worried by the fast bowling." But the description would be affected if the sentences were confused and unintelligible.

It is proposed in some tests to take into account the merits of the plan. It must be remembered that the plan must be subordinated to the general purpose of the composition. No doubt the best general rule in describing a cricket-match is to begin with the beginning. But in special circumstances it might be defensible to begin with a description

<sup>1</sup> The alternative title "objective tests" is used later in this book.

<sup>2</sup> For the sake of the foreign reader it may be stated that "l.b.w." and "M.C.C." are abbreviations for "leg before the wicket", and "Marylebone Cricket Club", respectively.

of some particularly striking episode and to work backwards and forwards from this point.

In the first investigation it is proposed to leave the investigators free to mark by impression and to compare their marks with those of other investigators acting in the same way. A discussion of the results may lead to more closely defined conditions of marking, and to a further comparison of marking by impression with marking by details (e.g., abundance of material, general structure, clearness of sentences, originality of expression).

8. The Class A tests were devised by Dr. P. B. Ballard and dealt with (i) synonyms (50 items); (ii) antonyms or "opposites" (50 items); (iii) spelling (30 items); (iv) punctuation and the use of capitals (40 sentences or sets of sentences); (v) combination of sentences (18 exercises).

9. The Class B tests were as follows :—

(1) Write a paper on the following subject as if for a school debating society :—

Has a person the right to spend his (or her) own money as he (or she) pleases ? Your object should be to state the arguments on both sides so that the society may clearly understand the position as you see it. You should illustrate your views by concrete examples.

(2) Describe a game of cricket (or hockey) as if you were writing for a French boy (or girl) who understands English, but has never seen the game played. Try to make your description lively as well as accurate, so that he (or she) may be interested in what you have to tell him (or her).

(3) You have seen your friend Bob (or May) off by liner to America. Write a description of the scene either (1) for a newspaper, omitting personal details, or (2) for a friend, Tom (or Lucy), who is also a friend of Bob (or May).

Before you begin the actual description, you are to state in a separate paragraph the precise object of your description, and for whom it is intended.

(4) You are interested in the career of a boy who is gifted but poor, and unable to go to a University without help. You are to write a letter as if to a wealthy friend (man or woman) asking if he (or she) will pay for the education of the boy, or lend him the money for it on the undertaking that he will pay the whole or a portion back at a later stage. You are to say in what way, if any, you are yourself helping the boy.

If you prefer it, you may write on behalf of a girl instead of a boy. You are to give names to the boy or girl on behalf of whom you write and to the person to whom you write.

10. It was obvious from preliminary correspondence that the scheme, asking for a period of  $2\frac{1}{2}$  hours for the writing of the essay, made too great demands on the schools to ensure the co-operation of any large number. A local education authority consented to allow schools under its control to be approached with a view to co-operation on condition that the source of any

scripts supplied should not be divulged, and that the chief inspector should see beforehand any report on the work which it was proposed to publish.

11. The following schools under this authority took part in the investigation :—

1 central boys' school <sup>1</sup>	...	School (a)
1 senior girls' school <sup>2</sup>	...	School (b)
1 senior boys' school <sup>2</sup>	...	School (c)
1 secondary girls' school	...	School (d)

In addition, the help was enlisted of two well-known public schools—one girls' school (e), and one boys' school (f). A third boys' public school (g) also offered its co-operation, but sent in its results too late for use. The essays in each school were, with one exception, marked (by impression) by one or more of the school staff; certain of them were also marked by outside examiners, including Dr. Ballard and the Director, and the essays of one school, besides being marked by a member of the staff, were marked by two members of the staff of another school.

12. The tabular statement below shows the details of this preliminary investigation (objective and essay-tests) :—

School	Age Group	Approx. No. of scripts for each of the 6 objective tests.	No. of essay- test scripts	Subject Number of essay- test	No. of individual markings for "sense" of essay-test	No. of individual marking for "sentences" of essay-test
(a)	12.6—13.6	40	37	3	2	2
(b)	12.6—13.6	40	39	3	6	6
(c)	12—13	40	36	2	2	1 <sup>3</sup>
(d)	12.6—13.6	40	30	2 <sup>4</sup>	3	3
(e)	13—15.6	51 <sup>5</sup>	43	4	4	4
		43 <sup>6</sup>				
(f)	13.6—15	40	35	1	1	3
<hr/>						
220						

13. After careful consideration of the essays of children of this age-group and of lower age-groups, and of the methods of

<sup>1</sup> A "central school" is attended by children from the age of 11+ who have been contributed by a number of neighbouring primary schools. They are of two kinds, selective and non-selective. As a rule the children enter the selective central schools by a competitive examination, generally the "special place" examination, and remain till they are 15+. The non-selective central schools come within the category of senior schools.

<sup>2</sup> A "senior school" is a public elementary school attended by children from the ages of 11+ to 14+ who have received their earlier education at a primary school, and have not been drafted off to some school of higher education, such as a secondary school, a selective central school, a trade school, a technical school, or a school of arts and crafts.

<sup>3</sup> The correction of the scripts presented special difficulties.

<sup>4</sup> Tests (i), (ii) and (iii) only.

<sup>5</sup> Tests (iv) and (v) only.

marking proposed by other investigators, the Director came to the conclusion that though it might be possible to attain some degree of consistency by confining attention to details of style ("sentences")—a method which Dr. J. H. Steel and Mr. Talman<sup>1</sup> seem to have pushed to its limits—consistency would be attained at the expense of validity, if the object of writing is to convey "sense" (see paras. 2–6 above).

14. He also came to the conclusion that the subjects were on the whole too difficult for pupils from the three schools below the standard of secondary schools, and that there was no possibility of marking them in any satisfactory way. The essays from the secondary schools were certainly better than those from central and senior schools, but it was noteworthy that the boys from a secondary school who dealt with subject (1) had little notion of giving arguments on two sides of a question. The descriptive subjects, Nos. (2), (3), and (4), set out in para. 9 above were far more fruitful.

15. The results of the investigation suggested, first, that the attainments in English of most school-children of from 12½ to 13½ were too elementary to be tested satisfactorily by an "essay" test, and, secondly, that in order to yield useful results any investigation of this kind should be made on pupils of a homogeneous group educated at schools of the same kind.

This led to the decision that the next investigation should be made on the work of boys and girls who were preparing for a School Certificate Examination and intended to take that examination within a year of the date of the essay-tests.<sup>2</sup>

<sup>1</sup> In their book on *The Marking of English Composition* (Nisbet & Co., 1936).

<sup>2</sup> The following examinations have been approved as First School Examinations by the Board of Education :—

- The School Certificate Examination of the Oxford and Cambridge Schools Examination Board ;
- The School Certificate Examination of the Oxford Delegacy for Local Examinations ;
- The School Certificate Examination of the Cambridge Local Examinations Syndicate ;
- The School Certificate Examination of the University of Bristol ;
- The School Certificate Examination of the University of Durham ;
- The General School Examination of the University of London ;
- The School Certificate Examination of the Northern Universities Joint Matriculation Board ;
- The School Certificate Examination of the Central Welsh Board.

On the origin and general structure of these examinations, see (1) the Government Report of which the full title is given in footnote <sup>(1)</sup> to para. 106 below ; (2) "Secondary School Examinations and the Curricula of Secondary Schools, with suggestions for reform," by Sir Philip Hartog, published by the National Union of Teachers, 1937 ; and (3) *A Conspectus of Examinations* (Macmillan & Co., 1937), published for the Committee.

## CHAPTER III.—INVESTIGATION ON THE MARKING OF ENGLISH ESSAYS, APPROXIMATELY OF SCHOOL CERTIFICATE STANDARD, AND THE RELATION OF THE MARKS TO THOSE AWARDED FOR CERTAIN OBJECTIVE TESTS

### NATURE OF THE SCHOOL POPULATION TESTED AND DESCRIPTION OF THE TESTS APPLIED

16. As a result of the experience of the preliminary investigation described above, it was decided that the essays investigated should all be written by pupils who were taking the School Certificate Examination within a year of the test. The schools selected comprised 13 boys' schools, 9 girls' schools, and 4 mixed schools ; and, with one exception, they were all secondary schools under Local Education Authorities. In a few instances schools presented parallel classes for the test. The total number of individuals who took part in the test was 1164, of whom 651 were boys and 513 were girls. The whole examination was spread, for the convenience of the schools, over several weeks ; and, owing to accidental causes, a number of the pupils were absent from some portion of the examination.

The tests were in many cases spread by the schools over a period of six months or more to avoid undue disturbance of school time-tables, clashing with the preparation of other examinations, etc.<sup>1</sup> 770 pupils in all, 441 boys and 329 girls did all the tests prescribed, namely six essays and four objective tests and an intelligence test.

#### *Average age of the pupils*

17. The average age of the pupils when the first test was taken was approximately 16½ years.

<sup>1</sup> No general announcement was made about the investigation and there is no reason to think that there was any leakage from school to school. Any such leakage would have shown itself by exceptionally high marks for the school to which it was conveyed.

*Subjects of the six essay-tests, directed and undirected, and time allowed*

18. In the Preliminary Investigation (see Class B tests, page 4 above) the pupils were required to write not only on a given subject but with a given audience and a given object in view. We have called such subjects "directed" subjects, as contrasted with the "undirected" subjects, mainly set at School Certificate examinations, complying with the general formula "Write anything on something for anybody," in which the only thing defined at the examination is the "something."

Each pupil was required to write an essay on each of the same six subjects. Of these three were "undirected" subjects which had been actually set at a School Certificate examination within a reasonable interval from the test, and three were "directed" subjects derived from the "undirected" subjects.

The following essay-subjects, three undirected and three directed, were selected for the investigation, with the approval of the main Committee :—

- |            |  |
|------------|--|
| Undirected | <ul style="list-style-type: none"> <li>A The right place of games in school life.</li> <li>B In praise of England.</li> <li>C A school speech-day.</li> </ul>  |
| Directed   | <ul style="list-style-type: none"> <li>A1 Should games be regarded as a duty or a pleasure ? Give pros and cons, as if you were arguing the case in a speech to a school debating society, and, if possible, come to a conclusion.</li> <li>B1 A French friend of your own age has told you in a letter that most French people consider their own country the best in the world, and asked you if English people think the same of England. You are to write in reply telling your friend in what respects you yourself think England is superior or inferior to any other country or countries.</li> <li>C1 Describe a school speech-day at which you have been present as if you were writing to a boy (or girl) of your class who has been prevented by illness from being present. Make your description as interesting and lively as you can.</li> </ul> |

Each pupil was required to write an essay on each subject. The subjects were therefore so chosen that the pupils might all reasonably be expected to have something of interest to say on each of them ; while such topics as "Gossip," "Gardens,"

and "Trees," all of which had been set in fairly recent School Certificate Examination papers, were avoided.

The time allowed for each essay was one hour, a period frequently allowed at School Certificate Examinations, and it was arranged that the undirected essay-subject should be set in each case before the corresponding directed essay-subject, and that the interval between the setting of the two related subjects should be in each case not less than three weeks, so as to avoid mere repetition.

### *Length of essays*

19. The average, minimum, and maximum number of words for the essays on the different subjects, as shown in a sample consisting of the essays of 20 candidates chosen at random, were as follows:-

Essay-subject	A	B	C	A1	B1	C1
Average number of words	559	570	577	559	550	570
Minimum	"	370	250	350	340	200
Maximum	"	770	1000	930	860	1080

### *Method of selection of the material for investigation<sup>1</sup>*

20. In order to bring the cost of the investigation within reasonable limits, it was decided that the final markings of the examiners (exclusive of the marking of trial-scripts; see paras. 27 and 30-33 below) should be limited to the 600 essays of 100 candidates, 50 boys and 50 girls, selected from the whole number in such a way as to provide an approximately "normal" distribution of merit (see para. 39 below). For this purpose the essays were marked in the first instance by the school-teachers of the pupils and their marks were furnished with the essays. They were allowed to use any method they pleased in this preliminary marking. The marks for each boy and each girl were then added together, and those of 50 boys and of 50 girls were selected so that the set of marks in both cases corresponded to what is known as the "normal" frequency curve. The Committee are much indebted for the helpfulness of the local authorities and of the headmasters and headmistresses and English teachers of the schools concerned.

It should be added that the marks of the teachers were not communicated to our examiners and that no other use was made of them except that described above for the preliminary selection of scripts.

<sup>1</sup> This subject is also dealt with in para. 39 below.

and "Trees," all of which had been set in fairly recent School Certificate Examination papers, were avoided.

The time allowed for each essay was one hour, a period frequently allowed at School Certificate Examinations, and it was arranged that the undirected essay-subject should be set in each case before the corresponding directed essay-subject, and that the interval between the setting of the two related subjects should be in each case not less than three weeks, so as to avoid mere repetition.

### *Length of essays*

19. The average, minimum, and maximum number of words for the essays on the different subjects, as shown in a sample consisting of the essays of 20 candidates chosen at random, were as follows:

Essay-subject	A	B	C	A1	B1	C1
Average number of words	559	570	577	559	550	570
Minimum "	370	250	350	340	200	320
Maximum "	770	1000	930	860	970	1080

### *Method of selection of the material for investigation<sup>1</sup>*

20. In order to bring the cost of the investigation within reasonable limits, it was decided that the final markings of the examiners (exclusive of the marking of trial-scripts; see paras. 27 and 30-33 below) should be limited to the 600 essays of 100 candidates, 50 boys and 50 girls, selected from the whole number in such a way as to provide an approximately "normal" distribution of merit (see para. 39 below). For this purpose the essays were marked in the first instance by the school-teachers of the pupils and their marks were furnished with the essays. They were allowed to use any method they pleased in this preliminary marking. The marks for each boy and each girl were then added together, and those of 50 boys and of 50 girls were selected so that the set of marks in both cases corresponded to what is known as the "normal" frequency curve. The Committee are much indebted for the helpfulness of the local authorities and of the headmasters and headmistresses and English teachers of the schools concerned.

It should be added that the marks of the teachers were not communicated to our examiners and that no other use was made of them except that described above for the preliminary selection

<sup>1</sup> This subject is also dealt with in para. 39 below.

INTERNATIONAL INSTITUTE EXAMINATIONS ENQUIRY

THE  
MARKING OF ENGLISH  
ESSAYS

A Report on an Investigation carried out by a Sub-Committee of the International Institute Examinations Enquiry Committee consisting of Sir PHILIP HARTOG (Chairman), Dr. P. B. BALLARD, Dr. P. GURNEY, Professor H. R. HAMLEY, and Dr. C. EBBLEWHITE SMITH

WITH A STATISTICAL REPORT BY  
DR. C. EBBLEWHITE SMITH

MACMILLAN AND CO., LIMITED  
ST. MARTIN'S STREET, LONDON.

1941

Price 3s. 6d. (By Post, 3s. 10d.)

and "Trees," all of which had been set in fairly recent School Certificate Examination papers, were avoided.

The time allowed for each essay was one hour, a period frequently allowed at School Certificate Examinations, and it was arranged that the undirected essay-subject should be set in each case before the corresponding directed essay-subject, and that the interval between the setting of the two related subjects should be in each case not less than three weeks, so as to avoid mere repetition.

#### *Length of essays*

19. The average, minimum, and maximum number of words for the essays on the different subjects, as shown in a sample consisting of the essays of 20 candidates chosen at random, were as follows :—

Essay-subject	A	B	C	A1	B1	C1
Average number of words	559	570	577	559	550	570
Minimum	"	370	250	350	340	200
Maximum	"	"	770	1000	930	860
						1080

#### *Method of selection of the material for investigation<sup>1</sup>*

20. In order to bring the cost of the investigation within reasonable limits, it was decided that the final markings of the examiners (exclusive of the marking of trial-scripts ; see paras. 27 and 30-33 below) should be limited to the 600 essays of 100 candidates, 50 boys and 50 girls, selected from the whole number in such a way as to provide an approximately "normal" distribution of merit (see para. 39 below). For this purpose the essays were marked in the first instance by the school-teachers of the pupils and their marks were furnished with the essays. They were allowed to use any method they pleased in this preliminary marking. The marks for each boy and each girl were then added together, and those of 50 boys and of 50 girls were selected so that the set of marks in both cases corresponded to what is known as the "normal" frequency curve. The Committee are much indebted for the helpfulness of the local authorities and of the headmasters and headmistresses and English teachers of the schools concerned.

It should be added that the marks of the teachers were not communicated to our examiners and that no other use was made of them except that described above for the preliminary selection of scripts.

<sup>1</sup> This subject is also dealt with in para. 39 below.

### *Mechanical reproduction of the scripts*

21. It was originally desired to reproduce the scripts by photography for simultaneous marking by the different examiners. This method of reproduction had been used in dealing with the scripts in French and Chemistry in earlier investigations.<sup>1</sup> The cost would, however, have been six times that actually incurred by making what are known as "Ormig" master-copies in typescript and rolling off the number of reproductions required;<sup>2</sup> while it was estimated that stencilled copies would have cost twice as much as the Ormig reproductions. It is of course true that a typed reproduction does not present to the examiner the idiosyncrasies of the pupil's handwriting, but in the present instance the handwritings were on the average so good and clear that few deductions of marks would have been made by any examiners on account of misunderstandings due to illegibility.

The typists were instructed faithfully to reproduce the mistakes in spelling and punctuation of the originals. The chief slips (very few in all) made by the typists were due to their involuntary corrections of mistakes. The reproductions were carefully compared with the originals and even if any slight oversight accidentally escaped attention in this process it would not significantly affect the validity of the investigation, since it was not the capacity of individual candidates but the consistency of the markings by different examiners that was being tested.

### *Nature of objective tests*

22. The objective tests used were as follows :—

Test 1	Group-Intelligence Test, series 33 of the National Institute of Industrial Psychology (193 items)
Test 2	Opposites (35 items)
Test 3	Spelling (40 items)
Test 4	Punctuation and the use of Capitals (58 items)
Test 5	Word-Meaning (26 items)

The objective tests were intentionally made short so as not to make too great demands on the time of the schools.

<sup>1</sup> See *An Examination of Examinations*, pp. 17-18; and *The Marks of Examiners*, paras. 67 and 98. Since this work, the cost of reproduction by photostatic and other photographic methods may have been reduced considerably.

<sup>2</sup> The firm manufacturing the "Ormig" machine in England is Block & Anderson, Ltd., of 30, Farringdon Street, E.C.4, to whom the Committee are indebted for their helpful assistance in obtaining good results. The capacity of a typist who is inexperienced in making "Ormig" master-copies cannot be assumed, but it is not difficult to train a typist for this purpose. A sufficiently heavy touch is essential.

Tests 2 to 5, reprinted in Appendix I (pp. 157-163 below), were devised by Dr. Ebblewhite Smith on the basis of those devised by Dr. Ballard for the Preliminary Investigation (see para. 8 above) and modified in the light of the results of that investigation. They were then finally settled after criticism by the Sub-Committee referred to in para. 24 below.

*"Opposites" Test.*—The questions used in the "Opposites" test were designed to measure the extent of vocabulary of the candidates; and the type of intelligence test question which tries to lure the candidate into underlining the wrong word was avoided.

*Spelling Test.*—The ordinary objective spelling test has two disadvantages :

(a) that the word is submitted to the testee without any context indicating its meaning, (b) that the correctly spelt word is accompanied by several misspellings which tend to confuse him.

The ordinary dictation test, on the other hand, has the disadvantage that it is wasteful of time, since the testee is obliged to write down many superfluous words (for instance, "the," "that," "in," "garden"), which 100 per cent. of the testees can certainly spell correctly.

The method used in the investigation has the advantage of the ordinary dictation test without its disadvantages. It will be seen from the test-paper that it consists of a number of printed sentences with a space in each left blank for the word which the testee is required to spell; that the supervisor in administering the test reads the completed sentence aloud twice at a moderate rate, and that the testee writes down only the "missing" word.

*Punctuation Test.*—The great difficulty with punctuation tests is to devise one which it is easy to score rapidly. In the test used this difficulty is got over by providing that in each line of the passages printed there shall be one mistake or omission and *one only*.

It is to be pointed out that the testee is not necessarily able to discover the mistake or omission by looking at each line in isolation. Thus in the first line of the first passage a parenthesis is omitted before the word "which," but it is only after reading the next line that the testee can discover that it is a parenthesis-mark, and not a comma, which has been omitted.

#### *Procedure adopted for the marking of the essay-scripts.*

*Personnel of the Board of Examiners and of the Sub-Committee of the main committee by whom the marking-scheme was devised*

23. The eight examiners (four men and four women) who actually marked the 600 essay-scripts (which were selected as explained in para. 39 below) had all had several years' experience in examining at the School Certificate standard, either as chief

examiners or as assistant examiners, or both, and had been recommended by persons acquainted with their work.<sup>1</sup>

In the Preliminary Investigation described above the examiners were unpaid, but in this investigation, as in the others carried out under the auspices of the main Committee, the examiners were paid in accordance with the usual scale adopted for the marking of scripts of the same kind and for attendance at examiners' meetings, or, in respect of some details, on a scale slightly higher. It might have been possible, though not easy, to secure for our purpose the voluntary help of a team of competent examiners; but marking by voluntary helpers would have been carried out under conditions different from those of a real examination. In an investigation of this kind it is to be remembered that the actual task of marking examination scripts is for most examiners wearisome, and the psychological condition of a person who is unpaid for performing such work is likely to be different from that of a person who is adequately paid.

24. The detailed plans for the investigation were settled by a Sub-Committee, working at the later stages in conjunction with the examiners, as will be explained below. The Sub-Committee consisted of Dr. P. B. Ballard, Professor H. R. Hamley, Dr. Percival Gurrey, Head of the Division of English in the University of London Institute of Education, and Dr. C. Ebblewhite Smith, Lecturer in the Postgraduate Division of the Institute,<sup>2</sup> with the Director as Chairman.

In planning the investigation the Sub-Committee had the benefit of the advice on statistical matters of Professor H. R. Hamley and Dr. Ebblewhite Smith, who had in mind the subsequent treatment of the results by the factor analysis method and the analysis of variance. The actual statistical analysis was carried out by Dr. Ebblewhite Smith, who was assisted by his wife, and, in the parts dealing with variance and factor analysis, by Dr. R. W. B. Jackson.<sup>3</sup> Several hundred correlation coefficients were calculated for the Committee by the Scientific Computing Service, Ltd., Bedford Square, London, W.C. Dr. Smith is responsible for nearly the whole of the final form of the Statistical Section of the Report.

25. The main object of the essay investigation was, as stated

<sup>1</sup> Their names have been treated as confidential. The individual examiners are referred to in this report as P, Q, R, etc., and always in the masculine gender.

<sup>2</sup> Recently appointed Associate-Professor of Education in the University of Manitoba, Canada.

<sup>3</sup> Now Research Assistant, Department of Educational Research, University of Toronto, Canada.

above, to ascertain whether "directed" essays, that is, essays written with a definite object and audience in view, would yield not only more valid results but more consistent scores than essays written in accordance with the "write anything about something for anybody" formula, in which only the "something" is defined.

Each of the examiners was informed of the arrangements which had been made with the schools, and was furnished, like the examiners for the preliminary investigation, with a copy of *Record or Message* and an offprint of the essay by the Director on composition in English in the volume of *Essays on Examinations* published by the Committee.<sup>1</sup> The Director also furnished them with an introductory and preliminary memorandum dealing not only with "sense" but suggesting categorics of details under which marks might be allotted.

26. In accordance with the traditional practice, the marking scheme was finally settled after the marking of trial-scripts by the examiners concerned, but the number of trial-scripts employed was exceptionally large (see para. 33 below), and the discussions, both oral and written, of the preliminary results, were exceptionally detailed.

Four meetings in all, each consisting of two sessions of approximately three hours, were held jointly by the examiners and the Sub-Committee: two on January 15th and February 12th, 1938, to discuss the marking of trial-scripts, one on March 31st, after 500 out of the 600 scripts had been marked, and a final meeting on July 16th, after all the 600 scripts had been marked in the ordinary way. At this final meeting the examiners discussed the preliminary results submitted by the statisticians, re-marked 14 clean Ormig copies of scripts that had been previously marked, and made suggestions for future marking schemes, on the basis of the work done. In the course of the work 16 memoranda or reports were furnished by the Director, containing either instructions for marking, or summaries of discussions at meetings, or of written reports of individual examiners, or of replies to questionnaires addressed to the examiners on various points. Only the major points in the procedure and decisions arrived at are set out in this report.

27. *Trial-scripts*.—Before the first meeting each examiner received 20 trial-scripts for preliminary marking on each essay-subject, 120 in all. Of each set of 20 trial-scripts 17 were original

<sup>1</sup> See *Essays on Examinations*: English composition at the School Certificate Examination; and the "Write anything about Something for Anybody" theory.

scripts, different for each examiner, and 3 were Ormig copies, the same for all the examiners. The object of these scripts was to familiarise the examiners with the material and to provide a common basis for the discussion of a marking scheme.

28. *Categories marked, and Scale of Marking.*—After the first discussion in connection with these trial-scripts, the examiners and the Sub-Committee decided that each essay should be marked for the following categories :—

(I) Sense ; (II) Spelling : (III) Punctuation, including Formal Paragraphing ; (IV) Grammar ; (V) Vocabulary ; (VI) Sentence Structure ; and (VII) General Impression ;

and that the mark for General Impression should be allotted *after* the other categories had been marked.

The question of the scale of marking was then discussed ; no one was (if we remember right) in favour of a 100 point scale. At the suggestion of Professor Hamley a 15 point scale was adopted in the following form : A+, A, A-, B+, B, B-, C+, C, C-, D+, D, D-, E+, E, E-. This may be considered as a refinement of the five-point scale. It has been used for many years in the University of London Institute of Education, and in London training college examinations.

It was also decided that marks were to be given by means of general estimates and not by means of negative marks for mistakes and positive marks for " felicities," although in practice the examiners marked for spelling by means of deductions.

29. *Median mark of scale to denote competency.*—It was decided that the median mark C should be regarded as designating competency, in other words, the possession of a " utilisable skill " in the writing of English. To avoid all misapprehension it is to be pointed out that C is the median mark of the scale, not the median mark of the candidates.<sup>1</sup>

30. *Marking of trial-scripts.*—Four additional Ormig scripts were marked at the first meeting ; and before the next meeting 12 new Ormig scripts were despatched to all the examiners, who were also asked to state whether in their opinion it was feasible in dealing with scripts of this kind to ask for marks for " Plan." The Director, in summarising the replies, said that, while he himself regarded the Plan as of the utmost importance in composition,

<sup>1</sup> The median mark of the candidates is a mark such that the number of candidates with a higher mark and of those with a lower mark is the same. It is to be noted that there can be no certainty that the median mark of the candidates denotes " a utilisable skill." If the capacity of the School Certificate candidates to write Greek prose were being tested it is highly improbable that the median mark would denote a utilisable skill.—P.J.H.

he had refrained from asking the pupils to append to their work a brief summary or plan for two reasons: (a) because so many of them were unaccustomed to make plans or summaries, (b) because the time allowed for writing (1 hour) was so short.

The 12 Ormig trial-scripts marked by the examiners were returned to each of them with the marks of all their colleagues, and with the request to revise them if on second reading they thought fit. But the revisions were few.

31. At the second meeting a detailed discussion took place on the 12 trial-scripts circulated after the first meeting, and on the revised marks which had been allotted by the examiners after seeing the marks of their colleagues; and three decisions were reached with regard to the marking-scheme:—

(i) To make the marking for "Plan" optional, it being left to examiners who found the column useful in arriving at a mark for General Impression to avail themselves of it.

(ii) In marking for Spelling, to allot marks only for the first two pages (in order to avoid penalising a pupil by reason of the length of his composition), and to mark down essays of less than two pages according to a settled scheme.

(iii) To provide a column for the marking of Literacy.

32. *Marking for Literacy.*—After long discussion it was found impossible to agree on a definition of Literacy (see paras. 106–9 below); but the column was inserted in order to see if the marks would disclose any general agreement in valuing under this head.

33. At this meeting each examiner also marked 13 Ormig trial-scripts; and, after scrutiny of the marks, the statisticians were of opinion that the marking of the 600 scripts should then proceed. It will be seen that before beginning to mark the final scripts each examiner had marked 149 trial-scripts of which 47 were Ormig scripts, identical for all examiners.<sup>1</sup>

34. *Marking-sheets; marking for Literacy.*—The marking-sheets for the 600 scripts were printed in the form appended (Appendix II), and for the use of the statisticians "assembly-sheets" of 36 pages for each essay-subject were drawn up showing side by side the marks allotted by each examiner for each category. In all, about 36,000 marks were allotted in this first part of the investigation. For the purpose of statistical calculation the literal marks E— to A+, were converted into the series 1 to 15, the median C,

<sup>1</sup> One examining authority has experimented by making the examiners correct 6 to 8 photographed scripts, identical for all examiners. It has published results in some subjects but not any record of the discrepancies of the marks in English, a significant omission. (See *The Marks of Examiners*, paras. 658–669.)

designating Competency, being rendered by 8. Only two examiners used the column for Plan to any great extent, and these marks have not been analysed statistically. The marks for Literacy were not inscribed numerically, but as either "Yes," "No," or "Doubtful." (See the chapter on Literacy, p. 80 below.)

### *Opinions and Recommendations of the Examiners*

35. In response to a written questionnaire the individual examiners expressed their opinions on various points which had arisen in the course of the investigation.

They were unanimous in thinking that the method of marking by categories was superior to the methods which they had previously used, though it demanded more time. The ordinary method, said one examiner, "requires an almost impossible balancing of the different qualities of an essay one against another"; and a second examiner added, "with this method the mind is kept more alert and balanced in judging the various elements of the composition."

They all agreed in regarding the category of "Sense" as of great value. The majority were decidedly of opinion that a "directed" essay-subject yielded an essay of better quality than the corresponding "undirected" essay-subject, and that it could be marked with greater confidence. Though there was a variation of opinion on these points, no examiner expressed the view that "directed" essays were harder to mark than "undirected." It was agreed that the quality of an essay, "directed" or "undirected," depended largely on the terms in which the subject was set.

There was a consensus of opinion that subject A1, a "directed" subject (asking for pros and cons in an argument), was too hard for the majority of the writers, and this view is borne out by the marks allotted. (See Table III, pp. 30-31 below.)

The Director, while of opinion that "directed" subjects should be set, expressed the view that under existing conditions it would be impossible to plough a candidate for irrelevancy, if he wrote a piece of decent English, because the training in English composition was in many schools so unsatisfactory. He himself would have no hesitation in ploughing a candidate for irrelevancy, if it did not, as at present, mean a plough on the examination as a whole, and if the candidates were allowed to take the subject of English composition separately on a subsequent occasion. The majority of the meeting were in agreement with this view. One examiner stated that at a certain examination assistant-

examiners were instructed to pass 60 per cent. of the candidates with "credit"; in such circumstances it would be impossible to plough a candidate merely because he did not comply with the rubric at the head of the examination paper.

35A. The whole question of categories was discussed again in great detail, and the question was considered whether it would be possible to assign a final mark by "weighting" the marks in the various categories and adding these together. No agreement was reached on this particular point; but the following recommendations were made:—

(1) That in the marking of essays the following categories and scale of marks should be used<sup>1</sup>:—

- (i) "Sense," with a 15-point scale A +, to E -, C denoting competency.
  - (ii) Spelling
  - (iii) Punctuation (including Formal Paragraphing)
  - (iv) Grammar and Syntax
  - (v) Accuracy of Vocabulary (the examiner bearing in mind only whether the words are used correctly or incorrectly)
  - (vi) Power of Expression, to cover (a) felicities of vocabulary, and (b) Sentence Structure
  - (vii) General Impression
- } with an 8-point scale, C to E -, C denoting competency<sup>1</sup>
- } with a 15-point scale, A + to E -, C denoting competency

(2) That the marks for "Sense" should be kept distinct from those in any other category.

(3) That, without any arithmetical weighting, the examiners might, as in the present investigation, form their estimates of "General Impression" after considering their awards under categories (ii) to (vi) inclusive.

35B. The examiners agreed generally that the marks which they had awarded under the heading "General Impression" corresponded to the final marks they would have awarded at a School Certificate examination, supposing that they had been free from the obligation to pass or give credit to a stated percentage of candidates.

<sup>1</sup> The Sub-Committee are not in complete agreement with the recommendations of the Examiners on the scales of marks to be used in marking categories (ii) to (v) above. On this point, see para. 169 (i) below.

*Re-marking of 120 scripts*

36. In order to investigate the "self-consistency" of the examiners the Sub-Committee decided to ask them to re-examine 20 scripts for each essay-subject (120 in all), after an interval of about five months from the completion of the original marking, and under precisely the same conditions in all respects as those of the first marking; and this was done. No previous notice of the request was given. The original numbers on the scripts were removed, and they were re-numbered so that it was impossible for them to be identified in any way with scripts previously marked. The actual choice of the 120 scripts to be re-marked was left to our statisticians. On the results of the re-marking see paras. 44, 96-105, and 117-122, below.

## CHAPTER IV.—STATISTICAL REPORT

By

DR. C. EBBLWHITE SMITH

NOTE.—*The general reader is advised to omit this chapter on a first reading.*<sup>1</sup>

37. The experimental methods used in this investigation and the way in which the "raw" data submitted to statistical analysis have been obtained are described in Part III above.

38. It is, however, desirable to recall here the main facts. 1164 pupils of secondary schools, boys and girls, all intending to take the School Certificate examination within twelve months of the beginning of the investigation, took part in it. The complete examination consisted of :—

(a) Six essays of an hour each on six different subjects, the same for each pupil ; (b) Intelligence group-test No. 33 of the National Institute of Industrial Psychology ; (c) four other new-type tests described in para. 22 above.

39. In order to avoid undue disturbances of school routine, the whole examination had to be spread by each school over several weeks, and owing to the unavoidable absences found to occur during such periods it was only possible to submit 770 pupils in all, 441 boys and 329 girls, to all the tests.<sup>2</sup>

We shall call these 770 individuals the "large sample." From this "large sample," 50 boys and 50 girls, whom we shall call the "small sample," were selected in the following way for the detailed markings of their essays on the same six subjects (three directed and three undirected—see para. 18 above).<sup>3</sup> It was

<sup>1</sup> This part of the book is intended to be self-contained, and was designed, within the limits available, to meet the needs both of statisticians and non-statisticians. It is realised, however, that while it contains statistical explanations superfluous for the professional, it may fall short in some places of the requirements of the non-statistician who must be referred to the standard textbooks on the subject.

<sup>2</sup> There is no reason to think that the pupils of any one school became acquainted with the tests from the pupils of any other school during the periods referred to.

<sup>3</sup> With reference to the selection of the small sample it should be stated that in a few cases the Ormig scripts on a given subject by some of the selected candidates were used as trial-scripts and were replaced in the actual investigation by Ormig scripts on the same subject written by other candidates who had obtained the same average mark from their teachers (see paras. 26-33). In all cases the scripts marked by all the different examiners were of course identical, so that the statistical results were not affected by this substitution.—P.J.H.

desired that the essays of the "small sample" should provide marks which would conform approximately to what is called "normal frequency" distribution; and for this purpose the class-teachers were asked to assign marks to the essays of all the 1164 pupils in accordance with any method that they deemed suitable, before the essays were transmitted to us; and they did so. The essays for the "small sample" were then chosen from the whole batch in such a way that the series of *total* marks for the six essays of the pupils selected, arranged in order of magnitude, conformed to the normal curve. This rough and ready method was sufficient for this particular purpose.<sup>1</sup>

40. The intelligence-test scores of the "large sample" were found to be distributed about an average (or mean) score of 133.44 with a standard deviation of 18.87 marks.<sup>2</sup> There were in the "large sample" differences both between the means and between the distributions of the intelligence-test scores of the boys and those of the intelligence-test scores of the girls. The mean score for the boys was 135.68 and the mean score for the girls was 130.51. The standard deviation of the scores of the boys was 18.33 and that of the scores of the girls was 17.73.

41. The distribution of the intelligence-test scores of the "small sample" was found to have a mean of 136.00 and a standard deviation of 18.61. These values are not significantly different from those of the mean and standard deviation of the intelligence-test scores of the "large sample." Sex differences were not considered in the "small sample."

42. Each of the 600 essays written by the candidates comprising the "small sample" was marked independently by eight examiners for seven categories: (I) Sense, (II) General Impression, (III) Spelling, (IV) Punctuation, (V) Grammar, (VI) Vocabulary, and (VII) Sentence Structure.

A few of the examiners gave marks for the Plan of each essay but the number of marks so awarded was not sufficient for statistical treatment (see paras. 30-1 above). Marks were also awarded by all the examiners for Literacy, but as these were not

<sup>1</sup> It will be remembered that the marks of the teachers played no further part in the investigation. They were not communicated at any stage to our examiners.

<sup>2</sup> The standard deviation is the usual measure of the spread of a distribution of marks. It is defined as the square root of the average of the squares of the deviations of marks from their mean. Consider the following example: 36, 34, 29, 27, 24 form a series of marks having a mean of 30. The deviations from this mean are + 6, + 4, - 1, - 3, - 6. The standard deviation  $\sigma$  is obtained by averaging the squares of these deviations and obtaining the square root, i.e.,

$$\sigma = \sqrt{\frac{36 + 16 + 1 + 9 + 36}{5}} = 4.43 \text{ marks.}$$

given in the same form as the marks for the other categories, it was decided to treat them separately. The main task of the statisticians was the analysis of  $341 = (6 \times 8 \times 7 + 5)$  sets of marks, or 34,100 marks altogether (exclusive of the marks for Literacy; see pp. 80-87 below), besides the 6720 marks yielded by the re-marking (see pp. 60-79 below).

43. The essay marks were based on a fifteen-point scale ranging from A+ to E-. For the purpose of statistical manipulation, these literal marks were converted into their numerical equivalents (1 to 15).

For the different objective tests the maximum marks were as follows :—

					<i>Maximum</i>
Test 1	Group Intelligence-test No. 33 of the National Institute of In- dustrial Psychology	...	...	...	193 marks
Test 2	A test of knowledge of verbal opposites	...	...	...	35 marks
Test 3	A test of spelling	...	...	...	40 marks
Test 4	A test of punctuation and the use of capitals	...	...	...	58 marks
Test 5	A test of knowledge of the meaning of words	...	...	...	26 marks

44. *Re-marking*.—For each essay-subject, representative samples, each consisting of 20 scripts, were chosen from the "small sample" and re-marked independently by each examiner for each category (see para. 36 above). This provided a further  $(6 \times 20 \times 8 \times 7) = 6720$  marks exclusive of the marks for Literacy.<sup>1</sup>

44A. In the text and in our tables, the following numbers will be used to designate the different categories :—

Sense	(I)	Grammar	(V)
General Impression	(II)	Vocabulary	(VI)
Spelling	(III)	Sentence	
Punctuation, in- cluding Formal		Structure	(VII)
Paragraphing	(IV)		

The several examiners will be designated by the letters P, Q, R, S, T, U, V, W.

The three "undirected" essay-subjects are called A, B, C. The three "directed" essay-subjects are called A1, B1, C1.

<sup>1</sup> The question of marks for Literacy is dealt with in Chapter V, pp. 80-92.

45. Tables I and II (1) to II (6) show representative samples of the original data in the form in which they were presented for analysis.<sup>1</sup>

TABLE I  
SAMPLE SCORES ON OBJECTIVE TESTS

Cand. No.	Test (1) (Intelligence- Test) (Max. 193)	Test (2) (Opposites) (Max. 35)	Test (3) (Spelling) (Max. 40)	Test (4) (Punctuation and Capitals) (Max. 58)	Test (5) (Word- meanings) (Max. 26)
9	149	33	36	50	21
23	137	31	37	53	21
38	146	29	35	48	19
43	118	30	25	36	13
61	165	35	39	55	20
81	152	32	36	46	18
6	110	29	40	49	21
13	132	31	35	54	18
52	136	30	31	33	17
59	125	31	34	56	11
69	86	22	27	44	8
85	167	35	37	50	19
11	134	31	40	51	19
22	138	29	39	53	17
43	118	30	25	36	13
45	156	33	37	51	16
68	137	34	39	51	21
99	157	32	23	50	11
4	138	31	37	56	20
12	171	31	38	52	23
68	137	34	39	51	21
80	117	30	35	53	12
93	124	28	30	45	11
98	84	25	28	52	13
2	127	26	21	49	6
32	124	33	34	50	15
35	152	28	38	49	22
41	143	32	39	53	16
93	124	28	30	45	11
95	161	33	39	57	21
25	116	24	31	44	16
37	119	31	38	51	17
66	122	26	38	55	18
68	137	34	39	51	21
82	100	30	39	57	15
95	161	33	39	57	21

<sup>1</sup> The complete figures are omitted to save expense.

**TABLE II (1)**  
**ESSAY-SUBJECT A**

Category	Cand. No.	Marks awarded to Scripts by Examiners :—							
		P	Q	R	S	T	U	V	W
(I) Sense	9	7	2	11	5	6	11	10	5
	23	8	10	11	10	6	8	8	12
	38	7	9	7	7	3	9	6	7
	43	4	4	6	4	3	6	4	2
	61	12	11	12	13	11	12	12	12
(II) General Impression	81	11	12	13	7	10	12	10	13
	9	7	7	10	5	6	11	10	8
	23	9	10	8	10	6	9	8	11
	38	6	7	5	6	3	9	5	5
	43	2	3	5	2	3	6	4	2
(III) Spelling	61	10	11	11	12	11	12	12	11
	81	12	12	12	8	10	12	10	12
	9	12	13	12	13	7	12	10	8
	23	11	11	12	11	6	11	14	11
	38	8	9	12	4	5	9	7	5
(IV) Punctuation	43	6	5	6	2	2	6	5	3
	61	13	13	11	13	11	12	15	11
	81	13	11	9	11	8	14	11	8
	9	7	9	12	8	7	11	10	9
	23	8	11	9	12	7	9	8	8
(V) Grammar	38	7	8	8	8	4	10	5	7
	43	3	2	5	2	4	4	5	4
	61	9	11	12	11	8	13	12	11
	81	11	12	11	10	8	12	11	9
	9	8	9	12	8	6	12	10	8
(VI) Vocabulary	23	11	11	9	14	8	11	8	11
	38	6	6	6	7	5	11	5	6
	43	4	2	7	5	4	8	6	5
	61	8	11	11	12	8	14	12	9
	81	11	12	11	13	8	14	11	9
(VII) Sentence Structure	9	6	10	11	8	7	8	10	8
	23	9	11	10	11	8	10	8	10
	38	6	9	8	7	5	8	5	7
	43	5	7	8	7	5	7	6	7
	61	11	11	11	12	11	13	12	11
(VIII) Punctuation	81	13	12	12	9	9	14	11	12
	9	6	9	9	8	6	10	10	9
	23	8	10	8	10	6	10	8	12
	38	5	6	5	5	2	8	5	4
(IX) Spelling	43	2	4	5	5	4	7	5	4
	61	10	11	11	12	11	12	12	9
	81	12	12	12	10	9	13	11	12

TABLE II (2)  
ESSAY-SUBJECT A1

Category	Cand. No.	Marks awarded to Scripts by Examiners :—							
		P	Q	R	S	T	U	V	W
(I) Sense	6	5	4	5	6	5	10	9	6
	13	4	6	7	6	8	7	7	2
	52	9	6	8	8	7	8	7	6
	59	8	9	11	6	4	11	8	10
	69	3	8	5	3	5	9	8	12
	85	11	13	12	7	4	9	11	12
(II) General Impression	6	8	9	7	6	6	10	9	11
	13	5	7	6	6	7	7	7	6
	52	7	7	8	9	6	7	7	9
	59	7	9	9	6	4	10	7	10
	69	2	5	4	2	3	6	7	5
	85	11	13	12	7	5	10	11	11
(III) Spelling	6	11	11	12	11	7	12	9	12
	13	13	10	13	10	8	12	8	9
	52	12	9	12	10	6	10	7	10
	59	12	14	9	10	5	12	8	11
	69	6	4	7	2	3	5	6	5
	85	12	14	13	11	8	11	11	12
(IV) Punctuation	6	9	11	12	8	7	9	9	11
	13	4	5	6	5	7	5	8	7
	52	7	8	7	11	3	7	7	9
	59	11	11	8	7	6	11	8	12
	69	3	1	4	2	3	3	7	3
	85	11	13	13	11	8	12	11	11
(V) Grammar	6	11	9	11	8	11	12	9	12
	13	8	6	12	11	7	8	8	8
	52	7	7	8	10	6	7	7	9
	59	9	11	11	8	6	9	7	11
	69	7	5	5	7	5	11	6	5
	85	11	13	14	11	6	10	11	11
(VI) Vocabulary	6	11	10	8	8	10	8	8	12
	13	7	8	12	8	7	7	7	8
	52	8	8	9	10	8	7	7	11
	59	7	9	11	7	7	10	7	12
	69	4	8	7	2	5	4	8	9
	85	11	12	12	11	6	9	11	11
(VII) Sentence Structure	6	9	10	7	8	8	10	9	12
	13	5	7	6	7	6	7	7	7
	52	5	7	8	9	5	6	7	9
	59	7	9	8	8	5	11	7	9
	69	2	6	4	2	5	5	8	6
	85	12	14	12	8	6	11	11	11

TABLE II (3)  
ESSAY-SUBJECT B

Category	Cand. No.	Marks awarded to Scripts by					Examiners:—		
		P	Q	R	S	T	U	V	W
(I) Sense	11	9	11	10	10	10	10	9	13
	22	7	8	10	6	8	11	8	5
	43	4	5	5	5	6	6	5	2
	45	5	11	8	9	10	9	6	5
	68	11	10	11	9	12	12	11	12
	99	7	6	7	8	8	8	5	8
(II) General Impression	11	9	12	10	11	10	10	9	13
	22	7	10	10	7	7	11	8	7
	43	3	4	5	5	5	5	5	4
	45	5	10	9	9	8	11	6	7
	68	11	12	11	9	11	12	11	12
	99	5	6	7	7	5	7	5	5
(III) Spelling	11	14	14	13	14	11	13	9	14
	22	13	12	13	12	7	12	8	8
	43	5	5	4	2	2	4	5	2
	45	11	7	12	7	7	8	7	7
	68	15	14	12	14	14	14	10	14
	99	8	5	7	2	2	5	5	4
(IV) Punctuation	11	8	10	12	11	7	10	9	13
	22	8	5	12	5	7	13	7	7
	43	2	5	8	7	4	5	5	7
	45	5	11	11	10	7	11	7	12
	68	11	9	12	11	11	12	11	8
	99	5	7	7	8	7	7	8	6
(V) Grammar	11	7	9	7	8	8	10	7	11
	22	9	10	12	10	7	12	8	8
	43	5	2	8	8	5	8	5	8
	45	7	10	12	11	8	14	7	8
	68	11	12	12	11	11	11	11	12
	99	6	5	11	8	5	8	7	5
(VI) Vocabulary	11	12	12	11	11	11	12	9	14
	22	8	10	9	8	8	11	8	8
	43	4	9	7	8	5	5	5	7
	45	6	11	11	9	9	12	7	9
	68	8	12	11	9	14	12	11	12
	99	5	8	8	8	6	8	5	7
(VII) Sentence Structure	11	9	11	11	10	11	12	8	11
	22	7	10	10	8	8	11	8	7
	43	2	2	5	7	2	5	5	7
	45	5	10	11	10	8	12	7	7
	68	7	12	11	9	11	11	11	10
	99	6	7	7	7	5	7	7	5

**TABLE II (4)**  
**ESSAY-SUBJECT B1**

Category	Cand. No.	Marks awarded to Scripts by Examiners :—						
		P	Q	R	S	T	U	V
(I) Sense	4	5	9	11	8	7	7	8
	12	8	13	11	14	14	10	11
	68	11	11	13	11	12	12	12
	80	7	8	8	8	5	7	7
	93	5	8	6	7	5	5	7
	98	2	5	6	5	7	2	8
(II) General Impression	4	6	11	11	9	8	9	9
	12	11	13	11	14	13	11	11
	68	9	10	13	11	12	13	12
	80	7	8	7	8	5	7	8
	93	6	8	6	7	4	5	7
	98	6	11	6	6	7	5	8
(III) Spelling	4	12	14	13	12	8	10	9
	12	12	14	12	14	14	13	11
	68	14	14	13	15	11	13	12
	80	13	13	12	12	7	9	7
	93	11	9	7	8	4	5	7
	98	12	14	11	11	11	9	8
(IV) Punctuation	4	6	11	13	13	8	11	9
	12	11	13	12	14	14	13	12
	68	11	11	13	12	11	13	11
	80	7	10	11	11	7	7	12
	93	8	8	6	8	5	8	8
	98	7	11	11	11	7	12	8
(V) Grammar	4	7	11	13	14	8	11	9
	12	8	13	12	13	14	11	11
	68	8	10	13	11	11	14	12
	80	8	8	11	7	7	7	12
	93	6	8	11	8	4	8	7
	98	6	11	8	7	7	11	8
(VI) Vocabulary	4	7	12	11	11	8	10	9
	12	9	14	11	13	14	11	9
	68	8	11	13	10	11	13	11
	80	6	8	8	8	5	8	12
	93	6	8	8	8	5	5	8
	98	7	13	8	12	12	11	9
(VII) Sentence Structure	4	6	11	11	10	8	10	9
	12	10	13	9	14	14	11	8
	68	8	9	13	9	10	11	11
	80	7	8	7	8	5	14	12
	93	7	8	7	8	5	8	8
	98	6	13	7	10	4	5	7

TABLE II (5)  
ESSAY-SUBJECT C

Category	Cand. No.	Marks awarded to Scripts by Examiners :—								
		P	Q	R	S	T	U	V	W	
(I) Sense	2	7	7	6	8	6	8	7	8	
	32	5	4	6	6	5	6	6	6	
	35	6	7	8	8	7	12	8	7	
	41	11	12	11	9	10	11	9	11	
	93	5	8	7	7	5	9	8	5	
	95	11	11	9	8	11	13	11	13	
(II) General Impression	2	7	7	5	8	7	9	7	8	
	32	6	5	7	8	5	7	6	7	
	35	8	7	8	8	8	11	8	8	
	41	11	12	10	10	10	10	9	12	
	93	5	8	6	7	5	8	7	6	
	95	11	11	9	8	11	11	10	12	
(III) Spelling	2	11	13	11	11	7	11	11	8	
	32	7	5	11	8	5	5	6	7	
	35	12	13	13	13	8	11	14	9	
	41	13	14	13	14	11	13	14	11	
	93	7	9	11	8	5	9	7	6	
	95	14	13	13	12	12	12	14	12	
(IV) Punctuation	2	7	8	11	8	7	12	9	8	
	32	7	7	14	8	7	8	8	7	
	35	9	7	14	11	7	11	9	7	
	41	10	12	12	10	8	9	9	13	
	93	5	9	6	5	6	12	7	6	
	95	11	12	13	8	10	11	11	12	
(V) Grammar	2	6	7	5	8	8	11	7	8	
	32	7	8	14	11	7	8	8	14	
	35	8	8	14	12	8	11	9	14	
	41	11	12	13	13	9	11	9	13	
	93	6	8	12	7	7	8	8	5	
	95	11	11	13	11	10	13	11	12	
(VI) Vocabulary	2	7	8	7	8	8	8	7	8	
	32	6	8	8	10	8	8	8	8	
	35	8	8	8	8	7	10	9	9	
	41	11	12	10	11	10	11	9	12	
	93	5	8	8	8	8	8	8	6	
	95	11	11	9	8	11	12	11	12	
(VII) Sentence Structure	2	8	8	6	8	8	8	8	8	
	32	6	8	8	9	8	8	8	8	
	35	7	8	8	9	8	10	9	8	
	41	11	12	10	10	10	9	9	11	
	93	5	8	6	7	7	7	8	7	
	95	10	11	9	8	10	12	11	13	

**TABLE II (6)**  
**ESSAY-SUBJECT Cl**

Category	Cand. No.	Marks awarded to Scripts by Examiners :—							
		P	Q	R	S	T	U	V	W
(I) Sense	25	13	11	11	10	10	12	8	11
	37	7	6	8	8	6	6	9	7
	66	6	8	7	7	6	8	8	7
	68	7	14	12	12	12	12	12	12
	82	7	9	9	5	6	7	7	7
(II) General Impression	95	11	11	9	9	7	11	11	8
	25	11	10	8	9	10	10	8	8
	37	8	8	8	8	6	7	9	8
	66	8	7	7	7	6	9	8	8
	68	8	14	11	12	12	12	12	12
(III) Spelling	82	8	8	9	6	6	8	7	7
	95	11	11	9	10	7	11	11	9
	25	11	10	9	9	7	9	7	5
	37	11	10	12	8	7	9	9	8
	66	11	14	13	11	11	11	8	8
(IV) Punctuation	68	14	14	14	14	14	14	12	12
	82	12	14	14	8	7	11	8	7
	95	13	14	12	14	8	11	11	9
	25	9	8	7	8	8	7	8	7
	37	8	10	12	11	7	8	9	7
(V) Grammar	66	9	6	12	8	5	12	8	6
	68	8	14	13	13	14	13	12	12
	82	8	9	12	8	7	11	8	8
	95	10	11	12	11	8	9	11	10
	25	11	11	12	12	9	12	8	8
(VI) Vocabulary	37	9	10	12	11	8	8	9	8
	66	8	7	12	8	6	11	8	8
	68	7	13	14	12	11	10	12	10
	82	8	7	13	8	6	11	8	8
	95	11	11	11	12	8	11	11	8
(VII) Sentence Structure	25	11	10	9	8	9	11	8	9
	37	8	8	8	8	6	8	9	8
	66	7	8	8	8	8	8	8	7
	68	7	14	11	12	12	12	12	12
	82	8	6	8	8	5	8	8	7
	95	9	11	8	10	7	11	11	9

*Measure of Agreement between Examiners*

46. The first question we shall ask of our data is "What measure of agreement exists between examiners on the marking of all the essays for all of the seven categories?" There are three aspects of this agreement, each of which will be considered separately:—

(i) Do examiners agree as to the general standard of the scripts, i.e. does Examiner X always tend to award the same marks to a particular script or batch of scripts as Examiner Y, or are his marks in general higher (or lower) than those awarded by Examiner Y?

(ii) How far do examiners agree as to the spread of their marks? It is frequently found, for example, that in examinations of the type here considered examiners refuse to use the highest and the lowest marks of the scale, thus reducing the effective spread of their marks. A measure of the extent to which the total scale is used by an examiner is the standard deviation of his marks.<sup>1</sup>

(iii) How far do examiners assign to scripts the same marks and rank them in the same order of excellence when they mark them for a second time without knowing what marks they assigned to them on the first occasion?

Questions (i) and (ii) are considered below; Question (iii) will be considered in the section on Re-marking (see pp. 60-79 below).

47. Although in conference the examiners had reached some general agreement as to the degree of merit denoted by each literal mark, it was evident from an inspection of their marks that some of the examiners regularly awarded higher marks than others. Table III gives the mean of the marks awarded to each essay-subject for each category by each of the examiners.

TABLE III -  
MEANS (FOR 100 SCRIPTS) OF EXAMINERS' MARKS  
(ON A SCALE 1 TO 15)

Category	Essay-subject	Examiner								Total	Mean
		P	Q	R	S	T	U	V	W		
(I) Sense	A	7.97	8.23	9.12	8.10	6.97	8.85	8.26	8.11	65.61	8.21
	A1	7.25	8.15	8.49	6.29	6.98	7.57	7.63	8.48	60.84	7.69
	B	7.14	8.98	8.59	8.19	8.69	8.88	8.12	8.55	67.14	8.30
	B1	7.65	9.13	9.02	8.95	8.22	8.86	8.52	8.48	68.83	8.63
	C	7.93	9.62	8.02	8.81	8.38	9.71	8.46	8.90	69.83	8.76
	C1	8.26	9.63	8.63	8.70	8.08	9.28	8.37	9.10	70.05	8.70

<sup>1</sup> In statistical terms, we have to compare the *means* (or averages) and the *standard deviations* (or spreads) of the marks awarded by the examiners and to calculate the *correlation coefficients* (or correspondences) between the marks of the examiners, considered in pairs.

TABLE III—*continued*

Category	Essay-subject	Examiner								Total	Mean
		P	Q	R	S	T	U	V	W		
(II) General Impression	A	7.72	8.38	8.24	8.02	6.65	8.92	8.10	8.00	64.03	8.00
	A1	7.32	8.00	8.27	6.66	6.70	7.99	7.61	8.72	61.87	7.73
	B	7.01	9.15	8.22	8.02	7.98	9.11	8.09	8.29	65.90	8.24
	B1	7.77	9.20	8.15	8.93	7.67	9.20	8.43	8.46	67.90	8.40
	C	7.80	9.47	7.76	8.75	8.14	9.56	8.24	8.54	68.26	8.53
	C1	8.27	9.57	8.16	8.45	7.83	9.30	8.48	8.61	68.76	8.60
(III) Spelling	A	10.98	11.30	10.52	10.24	6.93	10.43	10.20	7.88	78.48	9.81
	A1	11.74	11.50	10.73	10.23	7.45	10.01	8.36	10.05	80.07	10.01
	B	11.48	10.92	10.65	9.49	7.46	9.97	8.26	8.79	77.02	9.63
	B1	12.05	11.87	10.44	10.74	7.62	10.58	8.50	8.68	80.48	10.06
	C	10.75	11.67	11.48	10.48	8.33	10.40	10.49	8.52	82.12	10.27
	C1	11.20	11.23	11.37	10.19	8.00	10.48	8.59	8.45	79.60	9.95
(IV) Punctuation	A	7.81	8.56	10.13	9.20	6.04	9.04	8.53	7.46	68.57	8.57
	A1	7.81	9.11	9.33	8.84	7.23	8.20	8.28	9.18	67.98	8.50
	B	7.10	9.26	9.61	8.72	7.79	9.13	8.22	8.29	68.12	8.62
	B1	7.35	9.35	9.55	8.99	7.54	9.41	8.59	7.97	68.75	8.59
	C	7.55	9.41	10.63	8.05	8.20	10.22	8.83	8.16	71.05	8.99
	C1	7.82	9.12	9.85	8.63	8.25	9.69	8.85	7.83	70.04	8.76
(V) Grammar	A	7.89	8.59	10.53	10.11	0.91	10.29	8.23	7.89	70.35	8.70
	A1	8.05	8.65	10.14	10.17	7.28	9.58	8.23	9.83	71.93	8.00
	B	7.53	9.02	9.79	9.67	7.85	10.50	8.18	8.82	71.36	8.92
	B1	8.10	9.36	10.31	10.45	7.82	10.81	8.54	8.57	73.96	9.25
	C	7.95	9.54	11.06	10.23	8.10	10.28	8.59	10.32	76.16	9.52
	C1	8.60	10.02	11.35	9.99	8.61	10.03	8.83	8.70	70.91	9.01
(VI) Vocabulary	A	7.04	9.05	9.03	8.60	7.21	9.10	8.31	8.62	67.82	8.48
	A1	8.29	9.05	9.28	8.80	7.43	9.07	8.13	9.81	69.86	8.73
	B	7.45	9.57	9.25	8.69	8.49	9.51	8.22	9.49	70.67	8.83
	B1	7.81	9.82	9.28	8.93	8.19	9.84	8.56	9.34	71.77	8.97
	C	8.05	9.70	8.46	9.07	8.96	9.91	8.91	9.32	72.38	9.05
	C1	8.12	9.77	8.82	8.78	8.59	9.81	8.96	9.10	71.95	8.99
(VII) Sentence Structure	A	7.23	8.25	8.16	8.19	6.70	8.98	8.33	8.02	63.86	7.98
	A1	7.04	8.52	8.31	8.31	6.78	8.73	8.21	8.76	64.66	8.08
	B	6.42	8.95	8.24	7.88	7.88	9.24	8.34	7.94	64.89	8.11
	B1	7.10	9.00	8.11	8.25	7.37	9.63	8.60	8.34	66.40	8.30
	C	7.48	9.28	7.95	8.09	8.22	9.63	8.82	8.49	68.56	8.67
	C1	7.83	9.36	8.13	8.30	8.16	9.37	8.97	8.64	68.76	8.60
Examiner		P	Q	R	S	T	U	V	W		
Total		344	73	396	91	391	16	375	58	324	76
Mean		8.21	9.45	9.31	8.04	7.73	9.54	8.50	8.65	8.33	50

In general, Examiner U gives marks having a high mean and Examiner T gives marks having a low mean for all categories and essay-subjects. This tendency for examiners to maintain their relative standards of marking persists throughout the whole investigation and manifests itself in Table III as a positive correlation between any two rows of the Table. Examiners give higher marks for Spelling than for other categories.

48. In most cases, for any category, the mean of the marks given for each directed essay does not differ significantly from the mean of the marks awarded for the corresponding undirected essay. Further, there is no significant difference between the mean of the marks awarded to any essay, for Sense (I) and the mean of the marks for General Impression (II).

49. Table IV below gives the standard deviation of each examiner's marks for each category in respect of each essay-subject.

TABLE IV  
STANDARD DEVIATION (FOR 100 SCORERS) OF EXAMINERS' MARKS  
(ON A SCALE 1 TO 15)

Category	Essay-subject	Examiner								Total Mean	
		P	Q	R	S	T	U	V	W		
(I) Sense	A	2.211	2.235	2.090	2.247	2.274	2.090	1.809	2.778	17.783	2.223
	A1	2.609	2.624	2.287	2.308	2.522	2.255	1.922	3.163	19.710	2.461
	B	2.332	2.061	2.219	1.875	2.410	2.160	1.951	3.186	18.206	2.276
	B1	2.889	2.493	2.074	2.360	2.697	2.320	1.868	3.051	19.725	2.469
	C	2.060	2.181	2.267	1.927	2.481	2.113	1.969	2.624	17.463	2.182
	C1	2.373	2.517	2.028	2.052	2.374	2.259	1.831	2.464	17.895	2.237
(II) General Impression	A	1.960	1.782	2.285	2.276	2.238	1.798	1.836	2.417	16.592	2.074
	A1	2.420	2.064	2.293	2.303	2.243	2.042	1.865	2.076	17.906	2.238
	B	2.236	1.977	2.161	2.014	2.260	1.902	1.898	2.636	17.143	2.143
	B1	2.293	2.160	2.174	2.389	2.490	2.096	1.790	2.472	17.773	2.222
	C	1.761	2.090	2.329	2.007	2.445	1.639	1.919	2.369	16.579	2.072
	C1	1.788	1.991	2.048	2.027	2.425	1.714	1.857	2.116	15.966	1.996
(III) Spelling	A	1.930	2.583	2.334	3.020	1.834	2.039	2.782	2.224	18.866	2.358
	A1	2.120	2.748	2.235	2.025	1.862	2.536	1.425	2.431	18.282	2.285
	B	2.052	2.520	2.071	3.518	2.119	2.394	1.555	2.643	18.941	2.368
	B1	1.774	2.288	2.405	3.651	2.024	2.016	1.652	2.469	18.270	2.285
	C	2.002	2.728	2.234	2.759	2.522	2.040	3.266	2.464	20.016	2.502
	C1	2.159	3.023	2.110	3.029	2.530	1.900	1.632	2.504	18.977	2.372
(IV) Punctuation	A	1.683	1.774	2.318	2.319	1.696	1.902	1.723	1.884	15.329	1.916
	A1	2.357	2.306	2.775	2.792	2.130	2.615	1.537	2.487	18.909	2.375
	B	2.022	2.120	2.258	2.168	1.951	2.025	1.665	2.334	17.143	2.143
	B1	1.982	2.260	2.598	2.762	2.128	2.150	1.644	2.269	17.793	2.224
	C	1.693	2.093	2.715	2.295	2.400	1.708	1.732	2.266	16.092	2.124
	C1	1.846	2.338	2.450	2.103	2.070	1.901	1.410	2.254	16.083	2.123
(V) Grammar	A	1.794	1.934	2.100	2.209	1.727	2.026	1.886	2.332	16.008	2.001
	A1	2.338	2.071	2.514	2.458	1.721	2.426	1.502	2.289	17.319	2.165
	B	1.857	2.193	2.487	2.030	1.757	1.921	1.705	2.295	16.247	2.031
	B1	1.916	2.278	2.235	2.578	2.051	1.617	1.705	2.070	16.450	2.056
	C	1.590	2.012	2.584	2.310	2.360	1.738	1.887	3.159	17.649	2.206
	C1	1.765	2.010	1.867	2.007	2.457	1.665	1.379	1.950	15.106	1.888
(VI) Vocabulary	A	1.984	1.526	1.921	1.330	1.883	2.010	1.837	2.034	14.534	1.817
	A1	2.094	1.571	1.855	1.881	1.663	2.273	1.573	1.932	14.842	1.856
	B	2.105	1.818	1.810	1.748	1.942	2.017	1.809	2.124	15.442	1.930
	B1	1.983	1.717	1.709	1.914	1.901	2.019	1.734	2.237	15.214	1.902
	C	1.878	1.712	1.830	1.663	2.284	1.484	1.638	2.209	14.698	1.837
	C1	1.802	1.865	1.734	1.546	2.311	1.620	1.303	1.836	14.026	1.753

TABLE IV—*continued*

Category	Essay-subject	Examiner						Total	Mean
		P	Q	R	S	T	U		
(VII)	A	2.009	1.997	2.185	1.347	2.133	1.913	1.839	2.379
	A1	2.429	2.056	2.381	2.013	2.157	2.584	1.519	2.616
	B	2.359	2.165	2.030	1.512	2.206	2.001	1.762	2.630
	Sentenco	B1	1.936	2.203	1.969	1.946	2.314	1.958	1.631
	Structuro	C	1.763	1.882	2.066	1.719	2.431	1.683	1.577
	C1	1.625	2.032	1.853	1.572	2.675	1.896	1.292	2.076
Examiner		P	Q	R	S	T	U	V	W
Total		85.899	90.101	91.895	93.011	92.745	85.143	74.077	101.084
Mean		2.045	2.145	2.188	2.215	2.208	2.027	1.784	2.407

Examination of Table IV shows that examiners differ in the spread of their marks, and that examiners who give a wide range of marks for one category tend to do so for all categories. The standard deviations of marks awarded in respect of category I for directed essays are, in general, slightly higher than those for undirected essays.

50. It was expected that the eight examiners would agree in the marks awarded to directed essays for Sense and General Impression more closely than in those awarded to the undirected essays. But, as will be shown, this proved not to be the case.

51. It is to be pointed out that the "raw scores," i.e. the marks actually awarded by the examiners, do not yield immediately the necessary data for ascertaining the degree of their agreement as to the relative merits of individual scripts. In order to ascertain this we have to eliminate the differences of idiosyncrasy of the examiners in regard to two points which do not depend on the inherent differences between candidates—

(i) the differences generally described as differences of "standard" or "leniency" measured by differences between the mean marks awarded by examiners to the whole number of scripts; and

(ii) the difference of "spread" of the marks as measured by the "standard deviations" (see para. 40, footnote 2, and para. 46 above).

Let us consider these two points separately. If we find that the mean marks of one Examiner, X, are 10 per cent. higher than those of another, Y, we can render the comparison of the judgments of individual scripts fairer by raising the marks of Y by 10 per cent. for each script. This process will leave the order of the scripts unchanged, though it will alter the absolute value of the marks assigned to them by Y.<sup>1</sup>

<sup>1</sup> It is to be noted that this correction is based on the hypothesis that Examiners X and Y are steady in their estimates and does not allow for random variations.—P.J.H.

We now consider the second point. Let us suppose that the means of the marks awarded by Examiner X and Examiner Y are the same, but that the ranges of their marks are different. Thus, for instance, taking our fifteen-point scale, suppose that the mean mark of both Examiners X and Y is C, but that while the marks of X range from E to A+, those of Y range only from D- to B+, we may then say that while there is no difference in "standard" or average "leniency" between the two examiners, there is a marked difference in the "spread" of their marks, which makes it impossible to compare their judgments in regard to any particular script without allowing for this difference of spread.

52. The method of converting "raw scores" into "standard scores" takes account of *both* the points that we have considered, the differences in mean marks (or leniency), and the differences in the spread of marks of the two examiners. The "raw scores" of the examiners are converted into "standard scores" by means of a simple formula.<sup>1</sup>

Table V gives for the first two categories of essay A the "standard scores" of the marks given in Table II (1).

TABLE V  
SAMPLE STANDARD SCORES FOR CATEGORY I (SENSE) AND CATEGORY II (GENERAL IMPRESSION) FOR ESSAY-SUBJECT A

Category	Cand. No.	Standard Score of Examiner							
		P	Q	R	S	T	U	V	W
(I) Senso	9	-0.44	-2.78	+0.88	-1.11	-0.43	+1.02	+0.95	-1.12
	23	+0.01	+0.79	+0.88	+0.68	-0.43	-0.40	-0.14	+1.40
	38	-0.44	+0.34	-0.99	-0.39	-1.74	+0.07	-1.24	-0.40
	43	-1.80	-1.89	-1.45	-1.47	-1.74	-1.36	-2.33	-2.20
	61	+1.82	+1.24	+1.34	+1.75	+1.77	+1.50	+2.05	+1.40
	81	+1.37	+1.68	+1.81	-0.39	+1.33	+1.50	+0.95	+1.76
(II) General Impression	9	-0.37	-0.77	+0.77	-1.31	-0.30	+1.16	+1.03	0.00
	23	+0.65	+0.91	-0.11	+0.86	-0.30	+0.04	-0.05	+1.24
	38	-0.88	-0.77	-1.42	-0.88	-1.70	+0.04	-1.68	-1.24
	43	-2.92	-3.02	-1.42	-2.62	-1.70	-1.62	-2.22	-2.48
	61	+1.16	+1.47	+1.21	+1.73	+2.02	+1.71	+2.11	+1.24
	81	+2.18	+2.03	+1.65	-0.01	+1.56	+1.71	+1.03	+1.66

It will be noted that some of the scores are positive and others negative, indicating that some are above the average and others below. It will also be noted that the standard score, positive or negative, is seldom as great as 3.00.

<sup>1</sup> To obtain "standard scores" from a set of marks, we first subtract the mean of all the marks from each mark in turn and then divide the results by the standard deviation. The formula is:

$$z = \frac{x - \bar{x}}{\sigma}$$

where  $x$  is the original mark or "raw score",  $\bar{x}$  is the mean,  $\sigma$  is the standard deviation of the set of marks, and  $z$  is the "standard score."

53. The next question to be considered is this : Do examiners agree more closely in the marks they give to essays for the category " Sense " which is capable of an agreed definition, than they do for the rather ill-defined category " General Impression. "

54. In order to supply the answer to this question, the standard scores were further examined. The eight standard scores awarded to a script for one category were plotted graphically as a *cumulative frequency curve* having "standard scores" as abscissæ, and the number of examiners who awarded a certain mark or less as ordinates. The procedure may be explained by the following example : the standard marks awarded by the eight examiners to script 34, essay-subject C, category " General Impression " were

P	Q	R	S	T	U	V	W
-0.11	-0.22	+0.01	-0.88	+0.76	+0.28	+0.43	+0.19

When these are arranged in order of magnitude we have :—

-0.88	-0.22	-0.11	+0.01	+0.19	+0.28	+0.43	+0.76
-------	-------	-------	-------	-------	-------	-------	-------

This may be interpreted as meaning that

one examiner gives a mark	-0.88
two examiners give ..	-0.22 or less
three examiners give ..	-0.11 or less, etc.

Hence, plotting standard scores as abscissæ and the number of examiners who give a mark as low as  $y$  as ordinates, we obtain the cumulative frequency curve shown in Figure I.

FIGURE I

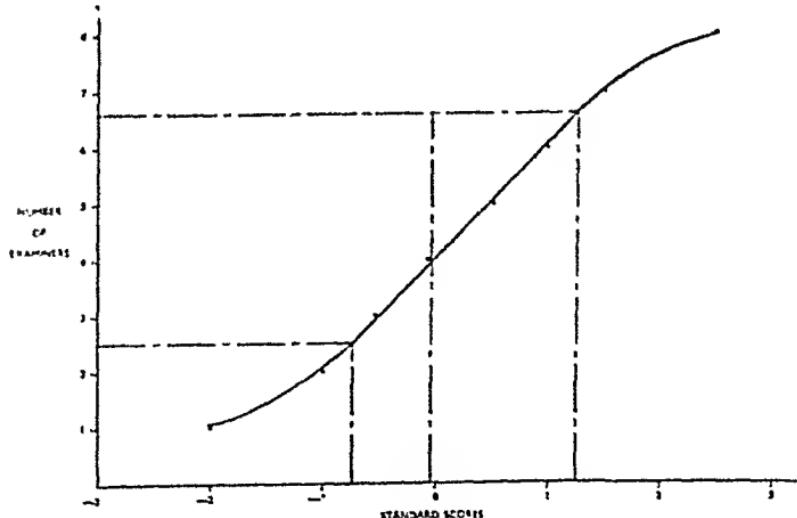


FIGURE 2

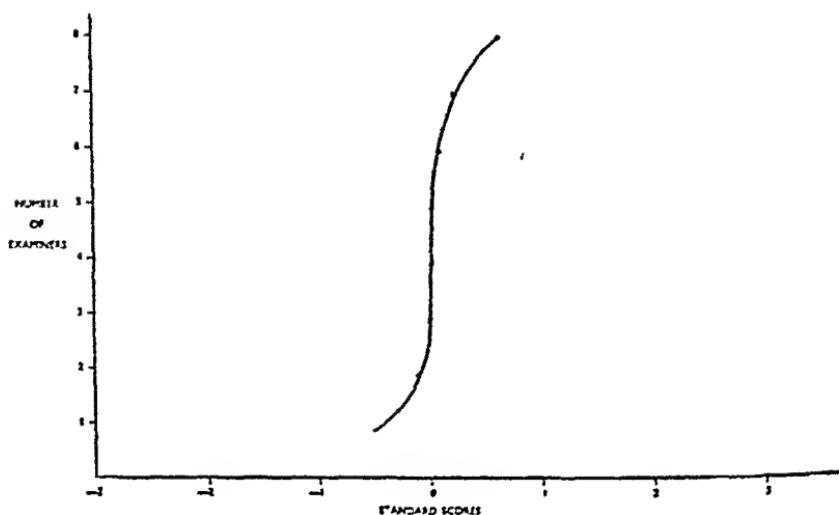
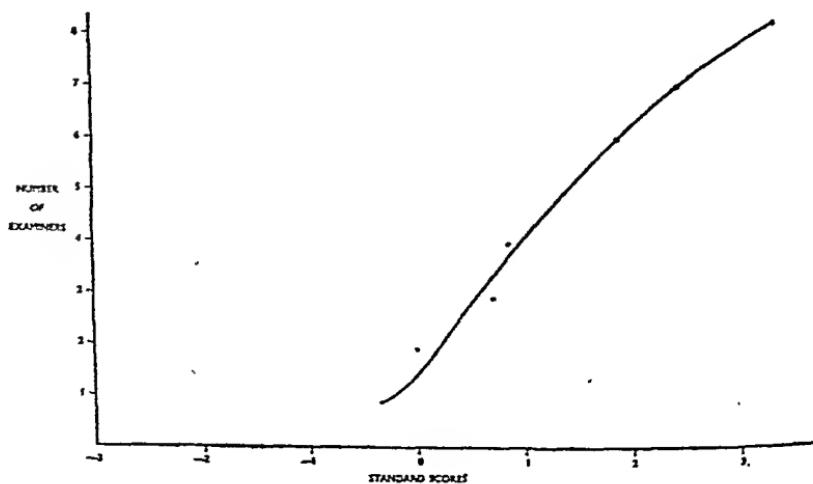


FIGURE 3



55. The curve here depicted is in the form of an S. This is the typical form for the cumulative frequency curve (ogive curve) derived from the normal frequency distribution (see para. 56 below).

It will be observed that one examiner rates the script 0.88 units below his average, and another 0.76 units above his average, while the other examiners give intermediate marks. If the examiners' marks had agreed more closely, the curve would have sloped more steeply; see Figure II. For a script for which there is less agreement, the curve will be more flat; see Figure III. Any measure of the steepness of the curve is therefore a measure of the agreement between the examiners. This is a better method of measuring the agreement between the examiners than measuring the total range of marks, because the range only takes into account two marks, the highest and the lowest. The difference between these two marks (the total range) is a measure of the extreme disagreement between examiners but is not a good measure of agreement, since the marks of one examiner may be outstanding while the others are in close agreement.

One measure of steepness is the interquartile range<sup>1</sup> which can be read directly from the graph.

56. The chief advantage of the graphical method, which is extremely laborious, is that individual abnormalities which do not show themselves in the usual statistical techniques are strikingly apparent in the graph. Because of the great amount of work involved, only two categories (Sense and General Impression) were considered. This involved the drawing of 1200 cumulative frequency curves (for 6 essay-subjects, 2 categories, 100 candidates) and the measurement of 1200 interquartile ranges. The great majority of the curves (96 per cent.) were found to be continuous and of approximately the same form as the ogive curve of the normal frequency distribution; see Figure I. This suggests to the statistician that in the majority of cases the differences between the awards of the eight examiners are due to the operation of a relatively large number of minor factors rather than to a small number of factors of great importance.<sup>2</sup>

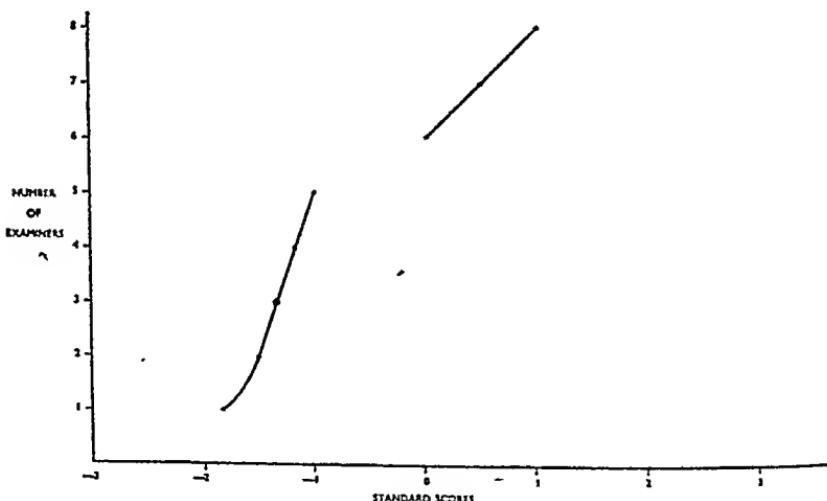
About 4 per cent. of the curves were discontinuous and a

<sup>1</sup> The interquartile range is the number of units measured along the base-line of the graph, between the point *a* and the point *b*, where *a* is that point below which one-quarter of the standard scores lie, and *b* is that point below which three-quarters of the standard scores lie.

<sup>2</sup> Normal distributions result when variations are due to a very large number of minor factors (see G. U. Yule, *Introduction to the Theory of Statistics*, Griffin, 1932, Ch. 15, para. 13, p. 306).

further examination of such cases showed that it was possible for examiners to hold widely different views on the excellence of the essay in question. For example: one candidate in a letter on the subject of Speech Day Proceedings, chose to adopt a facetious style and to use schoolboy slang. One group of examiners penalised the candidate heavily, regarding this interpretation of the subject as unsuitable, while others rated him fairly high because, in their opinion, he had fulfilled the requirements of the question. Figure IV illustrates quite distinctly this division of opinion among the examiners.

FIGURE 4



57. The cumulative frequency curves show, for different scripts, a wide range in the degree of agreement between the various examiners. The closest agreement was for candidate 70, essay-subject C1, category II (General Impression), where the interquartile range was only 0.10 units; and the greatest disagreement occurred for candidate 77, essay-subject C, category I (Sense), where the interquartile range was 2.10. The following Table gives the mean values of the interquartile ranges for each essay:—

	Essay-subject					
	A	B	C	A1	B1	C1
General Impression	.83	.75	.73	.84	.75	.78
Sense	.87	.78	.73	.95	.78	.84

Grouping the directed and undirected essay-subjects together, we have the following means :—

	A, B, and C	A1, B1, and C1	All Essay-subjects
General Impression	.770	.790	.780
Sense	.793	.857	.825

The only significant difference is that for Sense between essay-subject C and essay-subject A1. The comparatively high figure for Sense for essay-subject A1 suggests that, from the examiners' point of view, this subject was less suitable than the others. There is no significant difference between the average interquartile range of the examiners' marks for directed and that for undirected essays. Further, the somewhat unexpected result appears that the examiners do not show greater agreement when marking for a well-defined category such as Sense than they do when marking simply for General Impression.

58. If for the purpose of comparison we define "close agreement" of the examiners as corresponding to an interquartile range of less than 0·4, and "wide disagreement" as corresponding to an interquartile range of greater than 1·0, we find that in no instance were all the examiners either in close agreement or in wide disagreement in judging the six essays of a single candidate. There was a slight but noticeable tendency for examiners to agree more closely both when marking scripts for which they awarded very high marks and when marking scripts for which they awarded very low marks. This effect is a very familiar one to examiners and has been explained by the fact that for good scripts the range of possible marks is restricted at the upper end by the maximum mark and for poor scripts the range is similarly limited by the zero mark.

#### ANALYSIS OF VARIANCE<sup>1</sup>

59. Measurements of any human character differ from one individual to another. We may express these measurements as deviates from their mean and call the deviates variations. If we consider the variations of such a character as human stature, we shall find in general that they do not form a strictly homogeneous group. For example, two species may differ in their mean stature while there are wide variations from the mean within each species. Even within a single species it is always possible to find regional differences and genetic differences. Indeed, we can only be reasonably certain that variations form a homogeneous distribution when they are purely random, i.e. when they are caused by a large number of minor factors which

<sup>1</sup> For a Note on Analysis of Variance and Factor Analysis, see p. 149 below.

cannot be distinguished. Because it is so difficult in a research to plan an experiment yielding data having strictly homogeneous variation, it is customary to attempt to control the conditions of an experiment in various ways so that it may be possible to compare the variations of the data under different conditions. As a measure of the total variation of the data, we use the variance of the data, which is defined as the sum of the squares of the deviations of a series of values from their mean, divided by the corresponding number of "degrees of freedom."<sup>1</sup> In other words, the variance is the square of the standard deviation.

60. The 800 General Impression marks awarded to one essay-subject by our eight examiners constitute a group of measurements which may not be strictly homogeneous. It would seem to be of value to be able to differentiate the variation in this group of marks according to the various causes of its departures from homogeneity. Professor R. A. Fisher and his school have developed a method for analysing variance into its component causes.<sup>2</sup> We shall apply the analysis of variance technique to our data.

61. An inspection of the eight marks awarded by the eight examiners to any one script for any category reveals differences between the marks. In no instance have all the examiners awarded the same mark to a script (except in the case of marks for Literacy, where only a three-point scale was used). The differences are not solely due to differences in leniency of marking of the examiners, for if this were so, a simple adjustment of marks by addition or subtraction would equalise the marks. For, as we have previously suggested, if Examiner X marked consistently 10 per cent. higher than Examiner Y, then the differences between the marks of Examiner X and Examiner Y awarded to the same scripts could be removed by the addition of 10 per cent. to the marks awarded by Examiner Y. With our data, large differences still remain when this adjustment has been made. As may be seen from Table V, adjustment of marks

<sup>1</sup> The number of degrees of freedom (a phrase borrowed from mechanics) is found by subtracting from the number of measurements used the number of constraints or independent linear relations obtaining between them. Thus if  $N$  deviates are all measured from the same mean the number of degrees of freedom is  $(N - 1)$  since, when  $N - 1$  deviates are known the remaining one can be calculated, the algebraic sum of the deviates from the mean being zero. It can be shown mathematically that the number of degrees of freedom, rather than the number of deviates, is to be preferred in the estimation of standard deviations and variances. Where there is a large number of deviates,  $N$  is nearly equal to  $(N - 1)$  and  $N$  is often used as the divisor for convenience of calculation.

<sup>2</sup> It was originally applied to agricultural data but has since been employed in other branches of research. See R. A. Fisher, *Statistical Methods for Research Workers*, Oliver and Boyd, seventh edn., 1938.

both for differing leniencies of examiners and for different spreads of marks still leaves obvious differences between the eight marks awarded to one script. The questions we attempt to answer here are : How great are the differences due to unavoidable chance ? How great are the differences due to real differences of judgment between the examiners ?

62. If each one of a very large number of competent examiners, marking under the same conditions, awards a mark to the same script, then we shall not in general have complete agreement between the examiners. Practical experience leads us to assume that we may expect to find that there will be one most frequent mark which we shall call  $H$ , the remaining marks being distributed above and below  $H$ , the average mark, in accordance with the normal curve. Any single examiner's mark will usually differ from  $H$  for two reasons : (i) because the examiner will differ from his fellows in his estimate of the value of the script ; and (ii) because there will be unavoidable random (or chance) errors of marking. The mark  $H$  may be taken as an ideal measure of the "absolute merit" of the script.<sup>1</sup>

63. If it could be arranged that any *one* examiner should mark the same script many times, then, assuming that he neither remembers his previous mark nor is rendered careless through boredom, we may expect that more frequently than any other mark he will award a mark, say,  $G$ , but sometimes will award a higher mark, and sometimes a lower mark. His marks will be normally distributed about  $G$  as an average. We may consider  $G$  to be the ideal measure of that particular examiner's estimate of the merit of the script.

In general this will differ from the "absolute merit" of the script, and we may take the difference between the examiner's ideal mark  $G$  and the absolute merit  $H$  for the same script to be a measure of the examiner's idiosyncrasy  $C$ . Thus<sup>2</sup>

$$G - H = C \quad \dots \text{ (i)}$$

<sup>1</sup> The statistical methods used in this Report may perhaps be compared with those used by Dr. Rhodes and Professor Burt in the earlier investigations carried out under the auspices of the International Institute Examinations Enquiry and recorded in *The Marks of Examiners* (Macmillan & Co., 1936).

In the present Report three distinct methods of approach have been used : (a) the study of mean and standard deviations, which is similar to that used by Dr. Rhodes. The graphical analysis used in the study of marks for Senso and General Impression carries the study one step further ; (b) the study of variance, which was suggested by Dr. Rhodes in Chapter XII of *The Marks of Examiners* ; and (c) the method of factor-analysis, which was suggested by Professor Burt in his Memorandum on the Analysis of Examination Marks, in the book referred to above.—C.E.S.

<sup>2</sup> In this, and the equations which follow in this section, algebraic differences are to be understood.

64. When an examiner marks a script only once, his mark may, and indeed in most cases will, differ from his ideal mark for the script, being either greater or less than it by an amount  $E$ , which is called the random, or chance, error. If  $X$  is the mark awarded for one category by one examiner to a particular script, then

$$X - G = E \quad \dots \text{ (ii)}$$

Therefore

$$X - (C + H) = E$$

Or

$$X = H + C + E \quad \dots \text{ (iii)}$$

An examiner's mark may thus be considered to be made up of three parts :  $H$ , the absolute merit of the script ;  $C$ , the idiosyncrasy of the examiner, and  $E$ , the random error.

65. In what precedes we have been considering the marks awarded to one particular script, but have not considered whether this script is a good one or a bad one. For some scripts the absolute merit  $H$  will be high, for others it will be low. If the average value of  $H$  for all scripts on the same essay-subject is represented by  $A$ , then

$$H - A = B \quad \dots \text{ (iv)}$$

where  $B$  is the amount by which the absolute merit  $H$  of the script is above or below the average  $A$ .

$$\text{Therefore } X = A + B + C + E \quad \dots \text{ (v)}$$

Thus the mark  $X$  received by a candidate for his script is here defined in terms of four constituents : (1) the general average  $A$  ; (2) the candidate's deviation  $B$  from this general average ; (3) the idiosyncrasy  $C$  of the examiner ; (4) the random error  $E$  incidental to the marking.

If  $X_{st}$  is the mark given by a particular examiner  $s$  to a given script (or candidate)  $t$ , the above equation may be written more fully as :

$$X_{st} = A + B_t + C_s + E_{st} \quad \dots \text{ (vi)}$$

66. It is the aim of the examiners and examining boards to make  $C$  and  $E$  as small as possible, so that the final mark may be a trustworthy index of the real merit of the candidate's essay. To this end certain precautions are usually taken. The examiners meet to discuss the scale to be used in marking the essays, the categories for which marks are to be awarded, the standard expected at each "point" on the scale, and the possible idiosyncrasies of the examiners. A number of essays are then marked as trial-scripts, the results are fully discussed and the scheme of marking is modified, if necessary. In the present investigation a fifteen-point scale was used, since experience has shown that much finer scales are too fine for reliable judgment and coarser scales do not sufficiently differentiate the candidates. The precautions

$$\sigma_x = \sqrt{\sigma_B^2 + \sigma_\sigma^2 + \sigma_E^2};$$

$$\text{or } \sigma_x^2 = \sigma_B^2 + \sigma_\sigma^2 + \sigma_E^2 \quad \dots \text{(vii)}$$

where  $\sigma_B^2$  is the variance of the "absolute merit" marks  $H$ ;  $\sigma_\sigma^2$  is the variance of examiners due to their idiosyncrasies; and  $\sigma_E^2$  is the variance of the random errors incidental to marking. If  $(\sigma_\sigma^2 + \sigma_E^2)$  is small compared with  $\sigma_B^2$ , the mark  $X$  may be taken as a fair estimate of the candidate's real ability. If  $(\sigma_\sigma^2 + \sigma_E^2)$  is comparable to  $\sigma_B^2$ , or exceeds it, the marking may be considered to be unsatisfactory, since the mark awarded to the candidate is then determined by the idiosyncrasies of the examiners and by chance factors rather than by the ability of the candidate to write an essay.

70. Let us imagine that we have tabulated the marks given by each examiner to each script on a single category of an essay-subject, thus :—

TABLE VI

Candidates	Examiners							Total
	P	Q	R	S	T	U	V	
1								
2								
3								
4								
5								
—								
—								
—								
100								
Total								Sum

We wish to calculate (1) the variance *between* examiners, (2) the variance *between* candidates, and (3) the variance due to random error.

71. Let  $\bar{X}$  be the mean of all the 800 marks in Table VI. Let  $\bar{X}_m$  be the mean of the marks awarded to any candidate  $m$ , i.e. the mean of the marks in one row of the table. Let  $\bar{X}_n$  be the mean of the marks awarded by any examiner  $n$ , i.e. the mean of the marks in any one column of the table.

72. In order to calculate the total variance of the whole table, we find the sum of the squares of the 800 deviations from  $\bar{X}$ . This may be written :—

$$\sum_1^{800} (\bar{X}_i - \bar{X})^2 \quad \dots \text{(viii)}$$

where  $\bar{X}_n$  represents in turn each value in Table VI and  $\Sigma$ <sup>800</sup><sub>1</sub> means the sum of 800 such terms.

73. The sum of the squares of the deviates *between examiners* is calculated as follows: If we sum the marks in each column of Table VI, we shall have eight totals which are independent of the relative merits of the candidates and which differ only because examiners differ and because of random errors. We may divide each of these totals by 100 and thus derive the mean mark  $\bar{X}_n$  for that column. The difference between  $\bar{X}_n$  and  $\bar{X}$  represents an amount of variation (part of the total variation) which is independent of the candidates. But the deviation due to Examiner P is present in *each* of Examiner P's marks, that is, it occurs 100 times, so that the amount of the total variation due to the examiner and due to error but *not* due to the candidates in column P is  $100 (\bar{X}_n - \bar{X})$ .<sup>1</sup> Similar values may be derived for each of the eight columns. We may, therefore, write for the total variation independent of the candidates in the whole Table :

$$100 \sum_1^8 (\bar{X}_n - \bar{X})$$

or for the total "sum of squares" between examiners

$$100 \sum_1^8 (\bar{X}_n - \bar{X})^2 \quad \dots \text{ (ix)}$$

74. To find the "sum of squares" *between candidates* we must repeat this process with the means of rows. If each row in Table VI be summed and the means calculated, we have a series of values which are independent of the examiners and differ only because candidates differ and because of random errors. Let the mean value of row 1 be  $\bar{X}_1$ , then  $(\bar{X}_1 - \bar{X})$  represents the amount of variation (part of the total variation) which is independent of the examiners. This deviation, due to the ability of candidate 1 (and to random error) is present in each of his marks. Therefore it occurs eight times in the row and we may write the total deviation independent of examiners for the row as  $8 (\bar{X}_1 - \bar{X})$ . Similar values may be obtained for each of the 100 rows, so that for the whole table we may write :—

<sup>1</sup>  $\bar{X}_P$  is defined as the average mark of column P.

Total variation independent of the examiners

$$= 8 \sum_{1}^{100} (\bar{X}_m - \bar{X})$$

or the total "sum of squares" between candidates

$$= 8 \sum_{1}^{100} (\bar{X}_m - \bar{X})^2 \quad \dots (x)$$

75. If we divide the total "sum of squares" of equation (viii) by the number of degrees of freedom, 799, we have an estimate of  $\sigma_x^2$  (see para. 59 above). We call this  $V$ . In like manner it may be shown that the quotient  $V_1$ , obtained by dividing the "sum of squares" in equation (ix) by 7, is an estimate of

$$100 \sigma_o^2 + \sigma_x^2 \quad \dots (xi)^1$$

And, dividing the "sum of squares" of equation (x) by 99 we obtain  $V_2$ , an estimate of  $100\sigma_B^2 + \sigma_x^2$   $\dots (xii)$

It may also be shown that by subtracting the "sum of squares" between examiners plus the "sum of squares" between candidates from the total sum of squares, we have as remainder an estimate of the "sum of squares" of the random error. The "sum of squares" of the random error divided by the appropriate number of degrees of freedom, 693, i.e.  $(799 - 7 - 99)$ , is an estimate of  $V_3$ .

76. Part of our object in the analysis of variance is to compare the estimate of the variance of the error of marking with estimates both of the variance between candidates and of the variance

<sup>1</sup> For, consider the distribution of marks awarded to the 100 scripts by Examiner P. The spread of this distribution is due both to the spread of the merit of the candidates and to the random error. If we remove the candidate effect, then the situation is equivalent to one where Examiner P marks 100 candidates of equal merit, and the resulting marks are different only because of random error. The standard deviation of this random error is an estimate of  $\sigma_x$  based on 100 cases. The mean of these marks will therefore be subject to error, the extent of which is measured by  $\sigma_x/\sqrt{100}$ . Now, the means of the columns in Table VI are independent of the candidate effect, and, by the previous reasoning, each is subject to an error, the standard deviation of which is  $\sigma_x/\sqrt{100}$ . When, therefore, we calculate the standard deviation of these means, we obtain a value which is greater than the true standard deviation. We may write :—

$$( \text{Square of calculated standard deviation between examiners} ) = ( \text{square of true standard deviation between examiners} ) + \frac{\sigma_x^2}{100}$$

$$\text{or } 100( \text{square of calculated standard deviation between examiners} ) = 100 \sigma_o^2 + \sigma_x^2$$

between examiners. If, for example, the variance between candidates is the same as the variance of the error, then in general we may expect  $V_2/V_3 = 1$ . This would mean that differences in marks between candidates are not due to differences in ability between the candidates. In practice we should find that when the results, not only of this investigation but of all similar investigations that might be carried out, are brought together, the value of  $V_2/V_3$  would differ slightly from investigation to investigation, but the average value for all the investigations taken together would be 1. Professor Fisher has worked out a table for the distribution of  $V_2/V_3$ , and from his table it is possible to find the chances that the observed value of  $V_2/V_3$  has occurred by random fluctuation alone. Our procedure then is to apply Fisher's *Test of Significance*<sup>1</sup> to  $V_1/V_3$  and  $V_2/V_3$ .

77. The next step is to estimate the three standard deviations with which we are concerned ( $\sigma_s$ ,  $\sigma_c$ , and  $\sigma_e$ ) from the variances  $V_1$ ,  $V_2$ ,  $V_3$ , so that  $\sigma_s$ ,  $\sigma_c$ , and  $\sigma_e$  may be compared. The values of  $V_1$ ,  $V_2$ ,  $V_3$ ,  $V_1/V_3$ , and  $V_2/V_3$  are shown in Table VII.

TABLE VII

Category	Essay-subject	$V_1$	$V_2$	$V_3$	$V_1/V_3$	$V_2/V_3$
(I) Sense	A	40.82	25.46	2.14	19.06	11.80
	A1	58.48	28.48	3.08	19.01	9.26
	B	34.56	28.65	2.06	16.81	13.94
	B1	24.50	35.05	2.19	11.17	15.09
	C	44.68	27.65	1.60	27.95	17.30
	C1	28.97	27.11	1.96	14.76	13.81
(II) General Impression	A	42.28	23.43	1.68	25.09	13.90
	A1	64.06	27.09	1.98	32.40	13.70
	B	45.23	26.55	1.57	28.83	16.92
	B1	37.84	28.91	1.63	23.27	17.78
	C	47.93	25.60	1.39	34.48	18.41
	C1	35.67	22.54	1.43	24.08	15.79
(III) Spelling	A	240.48	32.80	1.90	126.43	17.25
	A1	215.11	31.54	1.76	122.12	17.91
	B	193.62	34.14	1.93	100.12	17.65
	B1	262.93	29.59	2.19	119.00	13.49
	C	151.28	35.70	2.30	65.65	15.49
	C1	187.36	32.12	2.17	86.43	14.82

<sup>1</sup> Fisher uses the function  $\frac{1}{2} \log_e (V_2/V_3)$  instead of  $V_2/V_3$  in his tables. This value he calls  $z$ . If, therefore, we wish to find from Fisher's tables the significance of  $V_1/V_3$  we must first calculate  $\frac{1}{2} \log_e (V_1/V_3)$ .

TABLE VII—*continued*

Category	Essay-subject	$V_1$	$V_2$	$V_3$	$V_1/V_3$	$V_2/V_3$
(IV) Punctuation	A	131.08	17.77	1.77	73.91	10.02
	A1	55.23	28.14	2.66	20.76	10.57
	B	69.14	21.69	2.28	30.27	9.50
	B1	76.51	22.85	2.57	29.79	8.90
	C	111.32	19.30	2.59	43.06	7.46
	C1	60.23	20.03	2.50	24.08	8.01
(V) Grammar	A	181.40	18.57	2.01	90.11	9.22
	A1	118.14	21.60	2.46	47.95	8.77
	B	106.58	19.01	2.12	50.23	8.96
	B1	133.47	20.02	2.12	62.92	9.44
	C	131.32	20.13	3.01	43.68	6.69
	C1	107.25	14.89	2.09	51.38	7.13
(VI) Vocabulary	A	43.47	17.06	1.44	30.24	11.87
	A1	56.22	15.66	1.80	31.24	8.70
	B	57.62	19.67	1.51	38.04	12.99
	B1	54.80	18.12	1.62	33.81	11.18
	C	37.30	16.98	1.55	24.07	10.96
	C1	32.62	14.49	1.56	20.86	9.27
(VII) Sentence Structure	A	49.63	20.56	1.66	29.89.	12.39
	A1	56.75	24.48.	2.33	24.40	10.52
	B	72.04	22.53	1.91	37.71	11.79
	B1	67.06	22.07	1.85	36.25	11.93
	C	48.66	18.07	1.73	28.08	10.42
	C1	34.37	17.12	1.80	19.13	9.53

78. Reference to the appropriate tables shows that in every case the variance due to differences between examiners is significant. That is, the marks given by different examiners to the same script do differ significantly. This is true for all categories and all essay-subjects.

79. Similarly, reference to Fisher's tables shows that the variance  $V_2$ , due to differences in the true merit of candidates' performances, is significant. This, of course, simply means that candidates do differ significantly in essay-writing capacity and that these differences are recognised by the examiners in the marks which they allot to the scripts. In other words, the differences between marks of candidates are not merely due to chance errors of marking or to differences in points of view of examiners.

TABLE VIII

Category	Essay-subject	Est. ${}^1\sigma_B$	Est. ${}^1\sigma_C$	Est. ${}^1\sigma_K$
(I) Sense	A	1.71	0.62	1.46
	A1	1.78	0.74	1.75
	B	1.82	0.57	1.43
	B1	2.03	0.47	1.48
	C	1.81	0.66	1.26
	C1	1.77	0.52	1.40
(II) General Impression	A	1.65	0.64	1.30
	A1	1.77	0.70	1.41
	B	1.77	0.66	1.25
	B1	1.85	0.60	1.28
	C	1.74	0.68	1.18
	C1	1.63	0.59	1.20
(III) Spelling	A	1.97	1.55	1.38
	A1	1.93	1.46	1.33
	B	2.01	1.39	1.39
	B1	1.85	1.62	1.48
	C	2.04	1.22	1.52
	C1	1.94	1.36	1.47
(IV) Punctuation	A	1.41	1.14	1.33
	A1	1.70	0.73	1.63
	B	1.56	0.82	1.51
	B1	1.59	0.86	2.60
	C	1.45	1.04	1.61
	C1	1.48	0.76	1.58
(V) Grammar	A	1.44	1.34	1.42
	A1	1.55	1.08	1.57
	B	1.45	1.02	1.46
	B1	1.50	1.15	1.46
	C	1.46	1.13	1.73
	C1	1.27	1.03	1.45
(VI) Vocabulary	A	1.40	0.65	1.20
	A1	1.32	0.74	1.34
	B	1.51	0.75	1.23
	B1	1.44	0.73	1.27
	C	1.39	0.60	1.25
	C1	1.27	0.56	1.25
(VII) Sentence Structure	A	1.54	0.69	1.29
	A1	1.66	0.74	1.53
	B	1.61	0.84	1.38
	B1	1.59	0.81	1.36
	C	1.43	0.69	1.32
	C1	1.38	0.57	1.34

<sup>1</sup>"Est." is here an abbreviation for "Estimates of."

80. Table VIII gives the values of our estimates of  $\sigma_c$ ,  $\sigma_e$  and  $\sigma_s$ . The results are interesting. In general, for any essay-subject  $\sigma_s$  is greater than either  $\sigma_c$  or  $\sigma_e$ . That is, in most cases, the merit of an essay appears to be the most important single factor contributing to the determination of the mark awarded to a script. The combined effect of other factors, however, may considerably exceed this and indeed does so in all categories other than Sense and General Impression.<sup>1</sup> It follows, therefore, that for categories III to VII a high mark awarded to a script by one examiner may be high as much because of random errors of marking and the bias peculiar to the examiner as because of the merit of the script in respect of that category.

81. Category (III), Spelling, and category (V), Grammar, have values of  $\sigma_c$  that are distinctly high. A reference to Table III, giving the average of the marks awarded by each examiner for each essay-subject and each category, shows that for category (III) Spelling, the examiners in general award higher marks than for the other categories, although this is not so in the case of Examiner T and is less marked in the cases of Examiners V and W. The average mark for Spelling awarded by Examiner T to all the essay-subjects is very much less than the corresponding marks awarded by Examiners P, Q, R, and S; in other categories Examiner T awards marks having averages of the same order as the average marks of the other examiners. It would seem, therefore, that, when marking for Spelling, Examiner T was doing so on a basis very different from that of the other examiners. This wide difference in point of view is probably the cause of the large value of  $\sigma_c$  for Spelling.

82. Similar marked differences between examiners are noticeable in the case of average marks for category (V), Grammar. In particular, it will be noted that Examiner T awards marks regularly lower than do the other examiners for this category and Examiner R awards marks regularly higher.

83. In the case of categories (VI) and (VII), Vocabulary and Sentence Structure,  $\sigma_c^2 + \sigma_e^2$  is approximately equal to  $\sigma_s^2$ .

84. There is no evidence from the figures in Table VIII to show that  $\sigma_c$  is significantly reduced in the case of directed

<sup>1</sup> By "the effect of a factor" Dr. Smith here means the contribution of the factor to the total variation as measured by the total variance. This is equivalent to measuring the effect by the figure commonly known as "variance" which is equivalent to the square of the standard deviation. Thus, taking the figures in the first row of Table VIII (for Sense), we have

$$1.71^2 \text{ is greater than } 0.62^2 + 1.46^2$$

(though 1.71 is less than 0.62 + 1.46). The reason is that we cannot add standard deviations but we can legitimately add variances.

essay-subjects, neither is there any evidence to show that the random errors of marking as estimated by  $\sigma_x$  are different in the case of the undirected essay-subject from those in the case of directed essay-subjects.

85. *Conclusions.*—We may summarise the results of our analysis as follows :—

(i) The marks awarded by different examiners to the same script do differ significantly. That is, where the marks given by two examiners to the same script differ, this difference is due in general, not only to random errors of marking, but also to real differences between examiners.

(ii) The absolute merit of the script does significantly affect the marks awarded to it. That is, in addition to the differences between examiners there is a significant measure of agreement between their marks.

(iii) Of the three components the absolute merit of the script has the greatest effect in determining the mark awarded, but the combined effects of the random error and the idiosyncrasy of the examiner are probably greater than that of the absolute merit for all categories except Sense and General Impression.

(iv) The categories Sense and General Impression were marked most satisfactorily ; for these categories the effect of the absolute merit of the script was very slightly larger than the combined effect of the random errors and the idiosyncrasies of the examiners. The categories Spelling, Punctuation, and Grammar were marked least satisfactorily.<sup>1</sup>

(v) In the categories Vocabulary and Sentence Structure, the combined effect of examiner differences and random errors is approximately equal to the effect on the marks of the absolute merit of the candidate.

(vi) Neither the examiner differences nor the random errors of marking are significantly less for directed than for undirected essay-subjects.

### FACTOR ANALYSIS<sup>2</sup>

86. In reading through an essay, the examiner receives a number of impressions, the totality of which fuse to give him

<sup>1</sup> It is to be remembered that all these categories were marked by impression and not by subtraction for errors from a maximum. This procedure was deliberately adopted in order to avoid the great expenditure of time which detailed marking of these categories would have involved. The object of the marking was to ensure that attention should be paid to each of these categories before the examiner marked for "General Impression" and not to require exact estimates for each category. See para. 28 above.—P.J.H.

<sup>2</sup> For a note on Analysis of Variance and Factor Analysis, see p. 149 below.

a general impression of the candidate's essay-writing ability. It is this that forms the basis of his General Impression mark for the script. Different examiners, each reading the same script, will all derive from it impressions of the handwriting of the candidate,<sup>1</sup> his spelling ability, his artistry, his command of language, his intelligence and so on, but they will differ among themselves as to the relative weights that they will place on each of these separate impressions in assigning the final mark. It is probable that in awarding their marks all examiners are influenced to some extent by every impression they receive from their reading of the script; but, while, for example, one examiner will place great stress on the "sense" of the essay, another will not do so, but will rather look for a high verbal ability on the part of the candidates, and he will not award a high mark, however convincingly a candidate has written, if he feels that the candidate has not shown evidence of a large vocabulary and a high degree of verbal fluency.

87. When marking for the various categories, the examiners were asked to award marks for only one impression. That is, the examiner was asked to give, as far as possible, zero weighting to all impressions other than the one defined by the category under consideration, so that when marking for Spelling, for example, he was asked to give no weight to Sense, Punctuation, Vocabulary, etc. The categories and the objective tests were chosen as representing the most important underlying influences likely to affect the examiner in making his judgment. By means of what is termed "Multiple Factor Analysis" it is possible to obtain an estimate of the relative importance of these influences in the marks awarded by the examiners for General Impression.

88. The technique of factor analysis which we have here employed is based primarily on the assumption that marks awarded by examiners may be expressed as the sum of a series of component marks variously weighted. Thus any mark  $X_{st}$  awarded by Examiner S to script  $t$  may be represented in the form :

$$X_{st} = a_1 x_1 + a_2 x_2 + a_3 x_3 \dots + a_n x_n \quad \dots \text{ (xiii)}$$

Here the numerical subscripts on the right-hand side of equation (xiii) represent those distinct categories or qualities in the essay which play *independently* an important rôle in producing a general impression of the script. Thus 1 may refer to Sense, 2 to Spelling, 3 to Grammar, etc. The letters  $a_1, a_2, a_3, \dots$ , etc.,

<sup>1</sup> In this particular investigation the handwriting factor was eliminated, as the essays were reproduced in typescript.

denote the weights a particular examiner  $s$  places on the categories 1, 2, 3, etc. They may be considered as the proportions of the total mark that an examiner assigns to each category. The values  $x_1, x_2, x_3$ , etc., are symbols denoting the marks awarded by the examiner to the different categories in each script.

89. Consider the following example : an examiner, in his General Impression marking, unconsciously awards, out of a possible fifteen, a maximum of ten marks for Sense, two marks for Spelling, two for Grammar, and fractions of a mark for the other categories, the total of all these fractions being unity. Suppose that the examiner judges the candidate to have for Sense an ability which he values as 60 per cent. (say 10 per cent. better than the average of his group); for Spelling 50 per cent., for Grammar 40 per cent.; and for the other categories other percentages. Then the mark awarded by the examiner will be :—

$$\begin{aligned} X_{st} &= 60\% \text{ of } 10 + 50\% \text{ of } 2 + 40\% \text{ of } 2 + \text{other small products} \\ &= 6 + 1 + 0.8 + \text{a fraction less than } 1 \\ &= 8 \text{ marks approximately} \end{aligned} \quad \dots \text{ (xiv)}$$

Another examiner marking the same script and unconsciously assigning different weights would, in all probability, award a different total mark to the script.

90. The factor analysis technique has been applied to our data in order to find an answer to the question : How many different values of  $a$  in equation (xiii) assume important proportions with our examiners and what qualities do these different  $a$ 's represent? <sup>1</sup> By analysing the values of the correlations between the marks of an examiner for each category and his marks for each other category, it would have been possible to have

<sup>1</sup> The fundamental equations of factor analysis are :—  
(See L. L. Thurstone, *The Vectors of Mind* (University of Chicago Press, 1935, Chap. I.)

$$\begin{aligned} a_1^2 + a_2^2 + a_3^2 \dots + a_n^2 &= 1 & \dots \text{ (i)} \\ a_1 b_1 + a_2 b_2 + a_3 b_3 \dots + a_n b_n &= r_{ab} & \dots \text{ (ii)} \end{aligned}$$

where  $a_1, a_2, a_3, \dots, a_n$  are the factor loadings of test  $a$   
 $b_1, b_2, b_3, \dots, b_n$ , are the factor loadings of test  $b$

$r_{ab}$  being the correlation between the two tests. It is unlikely that equation (ii) holds when we are correlating two sets of marks obtained by asking two examiners both to mark one set of scripts, since the weights that examiners give to the same category will differ, and also the impressions the examiners gain in respect of that category will differ. Equation (ii) above therefore becomes :

$$a_1 b_1 r_{11} + a_2 b_2 r_{22} + a_3 b_3 r_{33} \dots + a_n b_n r_{nn} = r_{ab} \quad \dots \text{ (iii)}$$

where  $r_{11}$  is the correlation between the two examiners' impressions in respect of factor 1, etc.

<sup>2</sup> In what follows correlation coefficients have been calculated by the Bravais-Pearson product-moment formula.

found the weights of various categories entering into his marks. This would have entailed forty-eight different analyses (there being eight examiners for each of six essays). In view of this, and remembering that a single examiner's marks are not as trustworthy as the average mark of a number of examiners, it was decided to average the marks of the eight examiners for each category and to use these average marks<sup>1</sup> in six analyses (one for each essay-subject).

91. Six tables of correlation coefficients (one for each essay-subject) have been prepared (Tables IX to XIV) each containing the intercorrelations of the marks (averaged for the eight examiners) for the seven categories, the marks for the intelligence-test and the marks of the four other objective tests. Table XV may be considered as common to all the Tables IX-XIV; and should form the lower right-hand part of each of them.

<sup>1</sup> Equation (xiii) then becomes :—

$$\bar{X}_{st} = \bar{a}_1 x'_1 + \bar{a}_2 x'_2 \dots + \bar{a}_n x'$$

where  $\bar{X}_{st}$  is the average of the eight marks awarded to one script

$a_1$  is the average of the weights awarded by the examiners to category I

$x'_1$  is an estimate of the true impression of category I, i.e. an estimate of the average impression of an infinite number of competent examiners.

## INTERCORRELATIONS FOR ESSAY-SUBJECT A

	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Cat. 6	Cat. 7	Obj. Test 1	Obj. Test 2	Obj. Test 3	Obj. Test 4	Obj. Test 5
Cat. 1	—	0.956	0.596	0.762	0.814	0.898	0.886	0.520	0.491	0.486	0.358	0.456
Cat. 2	—	0.714	0.870	0.921	0.945	0.964	0.556	0.536	0.561	0.445	0.511	
Cat. 3	—	0.712	0.717	0.657	0.697	0.697	0.390	0.380	0.625	0.418	0.474	
Cat. 4	—	—	0.846	0.789	0.851	0.494	0.483	0.562	0.502	0.499		
Cat. 5	—	—	—	0.884	0.934	0.468	0.508	0.573	0.466	0.500		
Cat. 6	—	—	—	—	0.948	0.518	0.561	0.530	0.410	0.547		
Cat. 7	—	—	—	—	—	0.525	0.546	0.550	0.439	0.493		

TABLE X<sup>1</sup>  
INTERCORRELATIONS FOR ESSAY-SUBJECT A1

	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Cat. 6	Cat. 7	Obj. Test 1	Obj. Test 2	Obj. Test 3	Obj. Test 4	Obj. Test 5
Cat. 1	—	0.932	0.546	0.712	0.700	0.828	0.810	0.405	0.462	0.390	0.348	0.286
Cat. 2	—	0.684	0.864	0.864	0.917	0.948	0.477	0.544	0.527	0.398	0.461	
Cat. 3	—	0.692	0.649	0.649	0.578	0.670	0.313	0.327	0.591	0.324	0.402	
Cat. 4	—	—	0.842	0.771	0.894	0.443	0.500	0.521	0.540	0.466		
Cat. 5	—	—	—	0.823	0.913	0.446	0.527	0.523	0.405	0.475		
Cat. 6	—	—	—	—	0.918	0.516	0.595	0.494	0.294	0.539		
Cat. 7	—	—	—	—	—	0.481	0.553	0.563	0.440	0.504		

<sup>1</sup> NOTE ON TABLES IX TO XVI

- (i) The essay-subjects A, A1, B, B1, C, CI are set out in para. 18 above.
- (ii) The categories referred to in these Tables as Cat. 1, Cat. 2, Cat. 3, etc., are the following —Category (I), *Senso*; Category (II), Spelling; Category (III), Punctuation, including Formal Paragraphing; Category (IV), Grammar; Category (V), Vocabulary; Category (VI), Sentence Structure; and Category (VII), General Impression (see para. 28 above).
- In the Tables, Arabic numerals have been substituted here in error for Roman numerals, but it is not thought that the substitution will produce any confusion in the mind of the reader.
- (iii) The objective Tests referred to are Test 1, Group-Intelligence Test; Test 2, Opposites; Test 3, Spelling; Test 4, Punctuation and the use of Capitals; and Test 5, Word-meanings.

TABLE XI<sup>1</sup>  
INTERCORRELATIONS FOR ESSAY-SUBJECT B

	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Cat. 6	Cat. 7	Obj. Test	Obj. Test	Obj. Test	Obj. Test
Cat. 1	—	0.911	0.563	0.390	0.730	0.805	0.782	0.522	0.535	0.441	0.129
Cat. 2	—	0.721	0.863	0.907	0.330	0.945	0.521	0.585	0.555	0.523	0.578
Cat. 3	—	0.692	0.654	0.683	0.676	0.369	0.313	0.655	0.119	0.462	—
Cat. 4	—	—	0.558	0.738	0.855	0.393	0.494	0.509	0.600	0.511	—
Cat. 5	—	—	—	0.849	0.916	0.412	0.533	0.537	0.518	0.572	—
Cat. 6	—	—	—	—	0.928	0.502	0.591	0.512	0.596	0.579	—
Cat. 7	—	—	—	—	—	0.486	0.581	0.505	0.493	0.526	—

TABLE XII<sup>1</sup>  
INTERCORRELATIONS FOR ESSAY-SUBJECT B1

	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Cat. 6	Cat. 7	Obj. Test	Obj. Test	Obj. Test	Obj. Test
Cat. 1	—	0.941	0.673	0.600	0.776	0.805	0.831	0.470	0.436	0.510	0.293
Cat. 2	—	0.764	0.851	0.913	0.932	0.953	0.507	0.462	0.357	0.400	0.493
Cat. 3	—	0.721	0.721	0.715	0.745	0.397	0.370	0.658	0.426	0.533	—
Cat. 4	—	—	0.844	0.773	0.875	0.411	0.379	0.500	0.197	0.433	—
Cat. 5	—	—	—	0.852	0.934	0.450	0.411	0.326	0.120	0.523	—
Cat. 6	—	—	—	—	0.928	0.159	0.131	0.188	0.355	0.318	—
Cat. 7	—	—	—	—	—	0.481	0.448	0.191	0.435	0.508	—

<sup>1</sup> See footnote (<sup>1</sup>) on page 55.

TABLE XIII<sup>1</sup>  
INTERCORRELATIONS FOR ESSAY-SUBJECT C

	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Cat. 6	Cat. 7	Obj. Test 1	Obj. Test 2	Obj. Test 3	Obj. Test 4	Obj. Test 5
Cat. 1	—	0.963	0.685	0.749	0.787	0.903	0.884	0.338	0.379	0.458	0.358	0.385
Cat. 2	—	0.778	0.850	0.875	0.950	0.945	0.390	0.471	0.542	0.446	0.472	
Cat. 3	—	0.772	0.741	0.723	0.742	0.358	0.406	0.639	0.455	0.504		
Cat. 4	—	—	0.833	0.833	0.879	0.355	0.499	0.533	0.608	0.494		
Cat. 5	—	—	—	0.851	0.895	0.409	0.513	0.544	0.503	0.544		
Cat. 6	—	—	—	—	0.952	0.398	0.505	0.510	0.406	0.497		
Cat. 7	—	—	—	—	—	0.386	0.520	0.502	0.460	0.490		

TABLE XIV<sup>1</sup>  
INTERCORRELATIONS FOR ESSAY-SUBJECT C1

	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Cat. 6	Cat. 7	Obj. Test 1	Obj. Test 2	Obj. Test 3	Obj. Test 4	Obj. Test 5
Cat. 1	—	0.612	0.645	0.776	0.884	0.896	0.338	0.406	0.448	0.258	0.410	
Cat. 2	—	0.749	0.787	0.877	0.931	0.969	0.412	0.486	0.537	0.389	0.528	
Cat. 3	—	0.656	0.722	0.634	0.688	0.407	0.447	0.710	0.514	0.566		
Cat. 4	—	—	0.770	0.710	0.800	0.346	0.465	0.488	0.478	0.460		
Cat. 5	—	—	—	0.810	0.886	0.387	0.433	0.526	0.384	0.520		
Cat. 6	—	—	—	—	0.945	0.380	0.470	0.445	0.300	0.525		
Cat. 7	—	—	—	—	—	0.412	0.496	0.495	0.367	0.524		

<sup>1</sup> See footnote (1) on page 65.

TABLE XV<sup>1</sup>  
INTERCORRELATIONS BETWEEN OBJECTIVE TESTS

Test	1	2	3	4	5
1	—	0·650	0·411	0·211	0·551
2		—	0·426	0·369	0·602
3			—	0·360	0·585
4				—	0·249
5					—

92. Each of the Tables above has been analysed by a method of factor analysis described by Professor Burt.<sup>2</sup>

It will be seen that for all of the six Tables of intercorrelations the results of the analysis are very similar. In each of these six Tables, one factor only is sufficient to account almost entirely for the intercorrelations between all seven categories. On removing the effect of this factor a group of small positive correlations remains between the objective test 1 (the intelligence-test) and objective tests 2, 3 and 5. It was further found that these remaining correlations may be accounted for on the supposition of a single group-factor common to all of these tests. The calculated weights of the two factors are given in Table XVI.

93. The factor pattern is difficult to explain. It is probable that the high value of the general factor in all the categories is due to what is known as a "halo" effect. That is, examiners when awarding a mark for other categories are unable to free their minds entirely from the General Impression, and their marks for all categories are really very strongly weighted with General Impression.<sup>3</sup>

94. The group-factor which occurs in objective tests 1, 2, 3 and 5 may tentatively be considered as a vocabulary factor having to do with a knowledge of the meaning of isolated words. It should be noted here that the intelligence-test was largely a verbal test.

<sup>1</sup> See footnote (1) on page 55.

<sup>2</sup> C. Burt, *Brit. J. Ed. Psych.*, Vol. VII (Part II), 1937.

<sup>3</sup> I had come to a similar conclusion from the inspection of the marks and discussions with examiners both during the preliminary period, and after the final markings. But as a matter of fact the examiners marked the scripts for "Sense" first, then for categories III to VII and Literacy, and finally for General Impression. The correlation between Sense and General Impression is, however, so high that the "halo" effect produced by marks for Sense would not be very different from that produced by General Impression.—P.J.H.

It has been assumed that the factors 1 and 2 retain the same meaning in the six analyses.—H.R.H.

## STATISTICAL REPORT

TABLE XVI<sup>1</sup>  
Factor Weightings on Loadings (See Paras. 87-90)

Essay	Factor	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Cat. 6	Cat. 7	Obj. Test 1	Obj. Test 2	Obj. Test 3	Obj. Test 4	Obj. Test 5	
		A	0.891	0.908	0.756	0.882	0.926	0.935	* 0.954	0.019	0.639	0.056	0.194	0.624
A1	1	0.827	0.954	0.711	0.896	0.898	0.012	0.959	0.587	0.056	0.634	0.175	0.602	0.439
	2	—	—	—	—	—	—	—	0.500	0.138	0.297	—	0.184	—
B	1	0.849	0.909	0.744	0.874	0.913	0.031	0.935	0.593	0.064	0.648	0.561	0.064	0.374
	2	—	—	—	—	—	—	—	0.517	0.147	0.180	—	—	—
B1	1	0.872	0.902	0.814	0.863	0.917	0.913	0.917	0.592	0.578	0.054	0.171	0.631	0.487
	2	—	—	—	—	—	—	—	0.498	0.551	0.108	—	—	—
C	1	0.873	0.952	0.820	0.899	0.913	0.933	0.916	0.514	0.612	0.613	0.531	0.617	0.452
	2	—	—	—	—	—	—	—	0.598	0.513	0.232	—	—	—
C1	1	0.859	0.960	0.806	0.818	0.892	0.899	0.947	0.534	0.013	0.653	0.462	0.651	0.104
	2	—	—	—	—	—	—	—	0.582	0.512	0.219	—	—	—

<sup>1</sup> See footnote (1) on page 65.

*95. Summary of Results.*

- (i) The observed correlations in the case of each essay-subject may be explained on the basis of one general factor.
- (ii) The general factor may best be described as a General Impression factor.
- (iii) The magnitudes of the factor are similar for directed and undirected essays.
- (iv) In the objective tests 1, 2, 3, and 5 we have found in addition to the general factor a group-factor which has been tentatively described as a verbal factor.

**THE RE-MARKING OF SCRIPTS (SEE PARAS. 36 AND 44 ABOVE)**

96. In order to test the "self-consistency" of the examiners it was decided to ask them, after an interval of five months, to re-mark representative samples of the scripts. For this purpose six samples, each consisting of twenty scripts, were chosen, one sample for each essay-subject. These were marked by the same eight examiners under the same instructions as before (see para. 36 above). Each examiner marked each script eight times; once each for the categories, Sense, General Impression, Spelling, Punctuation, Grammar, Vocabulary, Sentence Structure, and Literacy.

97. The Tables XVII (1) to XVII (6) give the marks awarded by the eight examiners, on re-marking, to the selected representative candidates whose marks at the first markings are given in Tables II (1) to II (6). It will be noted that there are large differences between marks awarded to the same script by the same examiner on the two occasions in all categories.

97A. As an index of the consistency of marking we have used the correlation coefficients obtained by correlating the marks awarded on the first occasion with those awarded on the second occasion by the same examiner. Correlation coefficients<sup>1</sup> have been calculated for each essay-subject and for each category (excluding Literacy, which was treated separately). The 336 ( $8 \times 6 \times 7$ ) correlation coefficients so obtained are shown in Table XVIII.

<sup>1</sup> As stated above, the correlation coefficients were calculated by the Bravais-Pearson product-moment formula.

TABLE XVII (1)  
ESSAY-SUBJECT A—RE-MARKING

Category	Cand. No.	Marks awarded to Scripts by Examiners							
		P	Q	R	S	T	U	V	W
(I) Sense	9	7	7	7	7	5	10	9	6
	23	7	10	12	9	10	11	11	12
	38	6	8	8	8	8	9	7	5
	43	4	6	4	7	7	6	6	2
	61	12	10	12	12	13	12	11	13
	81	11	12	11	9	11	12	9	12
(II) General Impression	9	7	8	6	7	6	10	9	6
	23	8	10	12	10	10	10	10	11
	38	6	7	6	8	7	8	7	6
	43	2	3	2	5	4	7	6	4
	61	10	10	12	12	12	11	11	13
	81	12	13	11	9	11	12	9	11
(III) Spelling	9	12	12	14	11	11	11	10	7
	23	11	12	15	10	11	11	10	7
	38	7	8	7	11	7	8	6	7
	43	6	7	6	5	5	8	3	5
	61	13	13	15	12	14	12	11	13
	81	13	11	10	11	10	14	7	11
(IV) Punctuation	9	7	10	8	8	8	11	9	7
	23	7	10	11	11	8	9	10	7
	38	6	8	11	8	8	9	7	6
	43	3	2	1	2	4	5	3	4
	61	9	9	12	12	12	11	11	12
	81	11	12	11	11	11	13	9	9
(V) Grammar	9	8	8	8	8	7	11	9	6
	23	9	9	12	12	11	11	10	12
	38	7	8	11	8	8	10	7	7
	43	4	5	5	5	4	11	6	6
	61	8	9	11	12	11	11	11	11
	81	11	10	12	11	11	14	9	11
(VI) Vocabulary	9	6	8	7	8	10	10	10	8
	23	9	11	12	9	10	11	10	11
	38	6	8	8	8	8	8	7	8
	43	5	5	5	7	7	8	6	8
	61	11	10	12	11	13	11	11	14
	81	13	12	11	8	11	12	9	12
(VII) Sentence Structure	9	6	8	6	8	7	12	9	5
	23	8	10	12	10	10	11	10	11
	38	5	6	6	7	5	8	7	5
	43	2	4	2	5	5	7	6	4
	61	10	9	12	11	13	11	11	12
	81	12	13	11	8	12	12	9	13

TABLE XVII (2)  
ESSAY-SUBJECT A1—RE-MARKING

Category	Cand. No.	Marks awarded to Scripts by Examiners							
		P	Q	R	S	T	U	V	W
(I) Sense	6	7	10	11	7	8	5	8	4
	13	5	8	6	7	9	8	7	4
	52	8	8	5	7	7	9	11	4
	59	8	7	9	7	11	10	9	9
	69	4	4	4	8	9	8	8	6
	85	9	8	11	8	9	8	10	13
(II) General Impression	6	8	9	9	7	9	7	8	8
	13	6	8	5	7	9	8	7	5
	52	8	7	5	7	7	9	11	5
	59	9	8	7	8	10	9	8	10
	69	4	5	2	7	7	7	7	4
	85	8	8	10	8	8	9	10	13
(III) Spelling	6	12	12	14	11	10	11	8	10
	13	11	13	14	11	10	11	8	8
	52	12	11	12	8	11	12	11	9
	59	12	11	14	9	10	12	8	8
	69	5	6	6	2	6	12	6	4
	85	11	12	12	10	8	11	10	12
(IV) Punctuation	6	7	11	13	8	12	11	8	11
	13	7	6	5	7	9	6	8	7
	52	9	8	6	8	7	10	11	7
	59	9	9	7	7	11	10	8	11
	69	4	3	1	2	6	4	6	3
	85	9	8	12	10	7	11	10	12
(V) Grammar	6	8	10	12	11	9	11	8	8
	13	7	8	9	8	9	11	7	8
	52	11	6	8	8	8	10	10	5
	59	8	7	11	7	11	11	8	10
	69	5	4	3	8	7	7	8	4
	85	8	8	14	11	8	11	10	14
(VI) Vocabulary	6	9	10	11	11	12	11	8	11
	13	7	8	8	8	10	10	7	9
	52	8	8	7	7	11	11	11	9
	59	9	8	7	8	10	9	8	11
	69	6	5	6	8	9	7	9	11
	85	8	8	10	11	8	11	10	11
(VII) Sentence Structure	6	8	9	11	10	9	8	7	9
	13	6	8	6	8	8	8	7	5
	52	8	6	6	8	7	9	10	5
	59	8	8	7	8	9	9	7	11
	69	2	5	2	9	7	8	7	5
	85	8	8	10	10	9	11	10	12

TABLE XVII (3)  
ESSAY-SUBJECT B—RE-MARKING

Category	Cand. No.	Marks awarded to Scripts by Examiners							
		P	Q	R	S	T	U	V	W
(I) Sense	11	9	11	9	10	11	12	11	12
	22	9	9	7	8	11	12	10	10
	43	4	4	5	7	7	5	6	3
	45	8	7	9	8	9	7	9	10
	68	11	8	11	9	14	12	10	14
	99	7	9	5	8	11	8	9	5
(II) General Impression	11	8	11	7	11	11	12	11	12
	22	8	9	7	8	11	12	10	11
	43	3	3	5	7	6	5	6	4
	45	8	7	9	9	9	8	9	10
	68	11	9	11	9	14	12	10	14
	99	6	8	5	8	9	9	9	5
(III) Spelling	11	14	14	14	14	12	14	11	13
	22	13	13	14	11	12	12	10	10
	43	5	2	5	4	5	5	6	4
	45	11	5	14	8	8	8	9	7
	68	14	14	13	14	14	13	10	14
	99	6	7	11	5	7	6	8	3
(IV) Punctuation	11	10	11	13	9	10	12	11	12
	22	7	11	11	8	12	11	10	12
	43	3	2	6	5	7	8	8	5
	45	7	8	11	11	9	9	9	8
	68	11	9	14	8	14	13	10	14
	99	6	8	7	8	7	7	9	3
(V) Grammar	11	9	9	6	11	11	12	11	12
	22	9	10	12	11	12	14	10	11
	43	5	2	10	8	7	8	8	7
	45	8	7	14	11	11	10	9	11
	68	11	8	13	11	14	13	10	14
	99	12	8	11	8	8	8	9	6
(VI) Vocabulary	11	9	11	11	12	12	13	12	14
	22	8	9	8	8	12	12	10	11
	43	4	5	6	8	9	7	8	6
	45	8	8	9	10	10	7	9	12
	68	9	9	11	10	14	12	10	14
	99	7	8	7	8	11	8	9	8
(VII) Sentence Structure	11	9	10	7	10	12	12	11	12
	22	7	10	7	8	11	11	10	10
	43	2	2	6	7	6	5	8	3
	45	8	8	9	11	9	9	9	10
	68	7	9	11	10	14	12	10	13
	99	5	8	5	8	8	7	9	5

TABLE XVII (4)  
ESSAY-SUBJECT B1—RE-MARKING

Category	Cand. No.	Marks awarded to Scripts by Examiners							
		P	Q	R	S	T	U	V	W
(I) Sense	4	6	10	10	9	9	8	10	12
	12	12	12	11	13	13	11	11	13
	68	11	11	13	9	13	10	9	13
	80	7	9	7	7	8	8	6	6
	93	5	8	5	8	9	6	9	4
	98	2	5	2	5	5	3	9	2
(II) General Impression	4	7	10	10	10	9	9	10	12
	12	11	12	11	13	13	11	11	13
	68	9	12	11	9	14	11	8	13
	80	7	9	6	8	8	8	6	7
	93	6	8	5	8	8	9	9	5
	98	6	9	2	6	5	5	9	6
(III) Spelling	4	12	14	15	11	11	11	11	12
	12	13	14	14	14	14	13	11	14
	68	14	14	15	12	14	13	8	13
	80	12	12	14	9	11	11	7	8
	93	10	11	7	8	7	7	8	3
	98	12	13	13	11	8	10	9	10
(IV) Punctuation	4	7	10	12	11	11	8	10	14
	12	11	13	12	13	14	12	10	14
	68	9	11	11	11	14	14	8	13
	80	7	10	9	8	10	9	7	7
	93	8	8	5	8	8	11	9	7
	98	7	11	6	10	8	9	9	7
(V) Grammar	4	10	10	14	13	11	11	10	13
	12	11	13	12	14	14	12	11	12
	68	8	10	14	8	14	11	7	13
	80	8	9	12	8	8	11	6	6
	93	5	8	11	10	8	11	9	7
	98	6	11	11	11	8	11	8	7
(VI) Vocabulary	4	8	10	11	11	11	10	11	12
	12	11	14	12	13	14	12	10	13
	68	8	12	13	9	14	10	8	13
	80	6	9	6	8	8	9	7	7
	93	5	8	6	8	8	8	9	8
	98	7	11	8	11	10	11	9	9
(VII) Sentence Structure	4	7	10	8	10	11	10	10	12
	12	10	12	11	13	14	11	11	13
	68	7	12	11	8	14	9	7	12
	80	7	9	7	8	8	10	7	7
	93	7	7	5	8	9	9	9	4
	98	6	11	6	10	8	11	9	7

TABLE XVII (5)  
ESSAY-SUBJECT C—RE-MARKING

Category	Cand. No.	Marks awarded to Scripts by Examiners							
		P	Q	R	S	T	U	V	W
(I) Sense	2	8	8	6	8	9	6	10	8
	32	6	5	7	8	6	4	6	6
	35	7	7	8	10	10	6	9	8
	41	12	12	8	10	12	10	11	11
	93	7	7	6	7	8	5	7	5
	95	12	11	9	11	12	11	11	13
(II) General Impression	2	8	8	6	8	9	9	10	7
	32	7	6	7	8	7	4	6	6
	35	8	8	8	10	10	7	9	8
	41	12	12	8	10	12	9	11	11
	93	7	6	5	7	7	5	7	5
	95	11	11	8	10	12	10	11	13
(III) Spelling	2	12	12	12	11	11	11	10	8
	32	8	11	10	7	7	4	6	6
	35	13	13	14	11	11	11	9	10
	41	14	13	14	11	11	12	11	12
	93	8	9	11	8	7	7	7	7
	95	14	14	14	11	12	12	11	12
(IV) Punctuation	2	10	8	12	11	8	10	10	7
	32	7	8	14	8	8	4	8	8
	35	11	9	12	11	10	10	9	7
	41	11	12	11	10	12	7	11	11
	93	7	6	6	5	8	8	7	4
	95	11	10	12	11	10	12	11	13
(V) Grammar	2	7	7	6	8	8	11	10	6
	32	7	8	14	8	8	8	7	8
	35	8	9	14	11	11	8	9	11
	41	11	12	14	12	11	11	11	12
	93	7	8	12	8	8	5	7	5
	95	11	10	14	11	11	11	12	12
(VI) Vocabulary	2	8	8	7	8	9	8	10	8
	32	6	8	9	10	9	8	8	9
	35	8	8	8	10	9	8	9	9
	41	12	12	7	9	11	11	11	11
	93	8	7	7	8	8	8	7	5
	95	10	11	8	8	11	11	12	12
(VII) Sentence Structure	2	7	7	7	8	8	9	10	8
	32	6	8	8	9	8	7	7	8
	35	7	8	8	9	10	8	9	7
	41	11	12	8	8	12	6	11	11
	93	7	6	6	8	8	5	7	4
	95	9	11	8	8	12	11	12	12

TABLE XVII (6)  
ESSAY-SUBJECT C1—RE-MARKING

Category	Cand. No.	Marks awarded to Scripts by Examiners							W
		P	Q	R	S	T	U	V	
(I) Sense	25	11	10	9	9	13	11	11	12
	37	8	9	7	8	9	6	10	8
	66	11	8	6	8	7	7	9	7
	68	12	12	12	10	14	14	11	14
	82	8	8	6	7	8	8	8	7
	95	11	11	11	8	13	11	10	12
(II) General Impression	25	9	9	7	8	13	10	11	9
	37	7	9	7	8	9	7	10	8
	66	9	8	6	8	7	8	9	7
	68	9	13	11	11	14	14	11	14
	82	7	9	6	7	8	9	8	7
	95	7	11	9	8	12	11	10	10
(III) Spelling	25	10	11	14	8	11	7	11	6
	37	11	12	14	8	9	8	10	6
	66	13	14	14	11	8	11	9	8
	68	14	14	14	13	14	14	11	14
	82	12	14	14	11	10	11	9	7
	95	13	14	14	11	11	12	10	12
(IV) Punctuation	25	6	8	7	7	14	9	11	8
	37	7	9	8	8	8	11	10	8
	66	8	8	7	8	8	11	9	8
	68	8	13	13	11	14	14	11	13
	82	7	10	13	8	8	11	9	8
	95	7	11	12	8	11	12	10	7
(V) Grammar	25	7	10	7	8	14	11	11	10
	37	8	9	14	8	9	10	10	11
	66	10	8	12	8	8	10	9	8
	68	8	13	12	11	12	13	11	14
	82	8	10	13	8	11	11	9	8
	95	8	11	11	11	12	11	10	11
(VI) Vocabulary	25	9	10	9	8	12	10	11	9
	37	7	9	8	8	9	8	10	10
	66	8	8	8	8	8	8	9	8
	68	11	13	12	11	14	14	11	14
	82	7	9	8	8	8	10	8	8
	95	7	11	8	8	12	11	10	7
(VII) Sentence Structure	25	7	9	8	8	13	11	11	7
	37	7	9	7	7	9	8	10	9
	66	8	8	6	8	7	5	8	7
	68	7	12	11	11	14	14	11	13
	82	6	8	7	8	8	11	8	7
	95	7	11	8	8	12	11	10	9

TABLE XVIII

CORRELATIONS BETWEEN FIRST AND SECOND MARKINGS OF 120 SCRIPTS  
(20 for each Essay-subject)

Category	Essay	Examiner							W
		P	Q	R	S	T	U	V	
(I) Sense	A	0.90	0.59	0.76	0.77	0.58	0.78	0.57	0.90
	A1	0.56	0.24	0.58	0.69	0.54	0.70	0.48	0.67
	B	0.73	0.49	0.74	0.77	0.74	0.79	0.54	0.54
	B1	0.85	0.82	0.73	0.75	0.76	0.83	0.29	0.51
	C	0.79	0.17	0.47	0.66	0.57	0.74	0.49	0.84
	C1	0.17	0.76	0.57	0.70	0.64	0.81	0.58	0.66
(II) General Impression	A	0.92	0.88	0.74	0.81	0.83	0.77	0.65	0.86
	A1	0.70	0.46	0.75	0.67	0.69	0.44	0.63	0.68
	B	0.81	0.71	0.80	0.83	0.80	0.66	0.44	0.79
	B1	0.95	0.74	0.83	0.78	0.77	0.73	0.13	0.87
	C	0.75	0.70	0.58	0.77	0.65	0.71	0.57	0.83
	C1	0.43	0.86	0.44	0.82	0.65	0.85	0.22	0.80
(III) Spelling	A	0.93	0.94	0.70	0.68	0.84	0.69	0.51	0.71
	A1	0.76	0.86	0.72	0.90	0.71	0.82	0.64	0.70
	B	0.84	0.88	0.76	0.92	0.85	0.78	0.59	0.89
	B1	0.89	0.92	0.86	0.83	0.72	0.63	0.31	0.86
	C	0.69	0.72	0.56	0.82	0.63	0.82	0.79	0.69
	C1	0.42	0.43	0.26	0.81	0.65	0.46	0.27	0.88
(IV) Punctuation	A	0.92	0.83	0.60	0.68	0.77	0.74	0.63	0.78
	A1	0.67	0.48	0.82	0.77	0.61	0.61	0.60	0.76
	B	0.80	0.48	0.88	0.63	0.63	0.58	0.41	0.51
	B1	0.89	0.82	0.81	0.86	0.75	0.67	-0.04	0.76
	C	0.68	0.46	0.80	0.73	0.17	0.40	0.61	0.83
	C1	0.46	0.66	0.57	0.51	0.49	0.29	0.55	0.61
(V) Grammar	A	0.87	0.79	0.46	0.73	0.77	0.31	0.56	0.76
	A1	0.25	0.50	0.62	0.49	0.36	0.17	0.59	0.60
	B	0.65	0.67	0.69	0.20	0.78	0.70	0.27	0.65
	B1	0.71	0.72	0.73	0.57	0.77	0.31	0.06	0.74
	C	0.67	0.48	0.49	0.68	0.42	0.45	0.45	0.53
	C1	0.14	0.69	0.61	0.72	0.19	0.31	0.55	0.69
(VI) Vocabulary	A	0.91	0.77	0.67	0.67	0.84	0.33	0.51	0.84
	A1	0.68	0.43	0.43	0.49	0.52	0.23	0.63	0.59
	B	0.78	0.46	0.87	0.73	0.62	0.46	0.54	0.65
	B1	0.89	0.80	0.67	0.79	0.78	0.65	0.11	0.80
	C	0.69	0.55	0.51	0.72	0.42	0.35	0.55	0.67
	C1	0.42	0.76	0.45	0.56	0.41	0.69	0.55	0.76
(VII) Sentence Structure	A	0.93	0.82	0.83	0.83	0.90	0.59	0.60	0.74
	A1	0.56	0.29	0.68	0.34	0.54	0.52	0.62	0.41
	B	0.65	0.79	0.72	0.68	0.74	0.58	0.46	0.59
	B1	0.91	0.71	0.75	0.87	0.63	0.41	0.07	0.82
	C	0.65	0.56	0.28	0.79	0.15	0.36	0.69	0.73
	C1	0.43	0.47	0.38	0.65	0.56	0.39	0.23	0.62

98. It will be observed that there is in the whole table a very wide range of values of correlation coefficients which vary from -0.04 (in one instance) to 0.95. Wide variations occur in the values for any one essay-subject, in the values for any one category, and even in the values for any one examiner. If we had chosen different samples of 20 scripts for each essay-subject for re-marking, the calculated correlation coefficients would have been slightly different. We may regard the correlation coefficients shown in Table XVIII as estimates of the true correlation (the average for an infinite number of samples between marks awarded on the two occasions).

It is possible to calculate the probability that an individual correlation coefficient in the table represents a measure of real agreement between the marks (i.e., a true correlation greater than zero). We shall accept as significant those correlation coefficients for which the chances are 100 to 1 that they represent a true correlation greater than zero. Only values of correlation coefficients greater than 0.56 satisfy this requirement. In other words, if we put 15 red balls marked 1 to 15 in one hat, and 15 black balls marked 1 to 15 in another hat, and take out a red ball and a black ball blindfold and note the numbers, replacing the balls and making twenty such selections, we should expect once in one hundred such sets of twenty choices, to find one value of the correlation between the red and the black numbers to be as high as 0.56. Of the 336 correlation coefficients in Table XVIII, 76 (i.e., over 22 per cent.) do not reach our criterion of significance.

99. The consistency of marking of any one examiner varies from essay-subject to essay-subject. Examiner P, for example, shows high consistency in his marks for essay-subject A in all categories, and low consistency in his marks for essay-subject C1 in all categories. The results of Examiner V show in all categories little consistency; indeed, in one instance (marks for Punctuation, essay-subject B1) the marks on the second occasion actually correlate *negatively* with the marks for the first occasion. Essay-subjects C and C1 appear to have been marked with least consistency in all categories. There is nothing to choose between the consistency of marking of the other four essay-subjects. In general, the categories Sense, General Impression, and Spelling have been marked most consistently, but there are in every category differences in consistency in respect both of examiners and of essay-subjects.

100. We cannot say from Table XVIII that the directed essays have been more consistently marked than the undirected essays. The differences of consistency, in general, between the

markings of the directed and undirected essays, are not significant, with the possible exception of those between the markings of essay-subjects A and A1. Here there is some slight evidence that the undirected essay-subject A has been marked more consistently than the directed essay-subject A1. Examiner P marks the undirected essays more consistently in all cases; this is not true for any of the other examiners.

101. We should expect the correlation between two markings by one examiner for the same category to be higher than the inter-correlations between the marks awarded by him for different categories. But this is not the case in our experiment; compare Tables IX to XIV with Table XVIII. In order to understand the reason for this the initial marks assigned to one hundred scripts for each category were examined in respect of each examiner separately. It was found that there was a strong tendency for examiners to award to a script the same marks for each category. This "halo" effect is particularly noticeable in the case of essay-subject C1 (see para. 93 above), for Examiner V, where identical marks were awarded to all seven categories for 48 of the 100 scripts, and for 93 of the 100 scripts identical marks were awarded to four or more categories! The "halo" effect between categories is much greater in the case of Examiner V than for any of the other examiners, yet the awards of all the other examiners show a similar tendency. The effect of this tendency on the part of the examiner is to produce an abnormally high correlation between his marks for the different categories.

102. The coefficient of correlation between the first and second markings is not of itself a complete measure of consistency.<sup>1</sup>

We must ask two further questions:—

- (i) Is the standard deviation of the marks the same for the first and second markings?
- (ii) Is the average mark the same for the first and second markings?

If, for example, an examiner on re-marking twenty scripts exaggerates the value of good scripts by awarding higher marks than at the first marking, and depreciates the value of poor scripts by giving lower marks than on the first marking, he may still rank the scripts in the same order of merit and so have perfectly correlated sets of marks yet no two marks awarded

<sup>1</sup> See R. W. B. Jackson, on the "Reliability of Mental Tests," *Brit. J. Psych.*, XXIX, Part 3, p. 267 (1939).

to the same script may be the same. Again, if his standard of severity changes from the first marking to the second, then the rank order of the marks awarded may be the same on the two occasions yet all the marks on the second marking may be lower, or all may be higher, than those on the first marking. Our results show that for the first and second marking of the twenty scripts there are both differences in standard of marking and differences in spread of marks awarded, irrespective of the correlation between the two sets of marks.

Consider, for example, the following results for 20 scripts marked by Examiner T.

#### ESSAY-SUBJECT B

	<i>Category IV, Punctuation</i>		<i>Category VI, Vocabulary</i>	
<i>Mean</i>	<i>1st Marking</i>	<i>2nd Marking</i>	<i>1st Marking</i>	<i>2nd Marking</i>
<i>Standard</i>	7.05	9.75	7.95	10.45
<i>Deviation</i>	1.47	2.24	1.91	1.67
<i>Correlation</i>	0.627		0.620	

There is here no significant difference between the correlation coefficients. In both cases the average mark awarded on the second marking is significantly greater than that on the first marking. The standard deviation of the second marking for Punctuation is significantly larger than that of the first marking. There is no significant difference between the standard deviations of the first and second markings for Vocabulary.

103. As we have pointed out above, while correlation-coefficients and standard deviations give valuable information on the statistical relations of two series of figures, they fail to give some of the most important facts needed by examiners and examining bodies, which can only be obtained by comparing the actual marks awarded. We have given above the actual marks for all the categories for the two markings of 6 samples of essays for each essay-subject. In Tables XIX (1) to XIX (6) below, we give the marks for General Impression, awarded by all the examiners at the first and second markings of the whole set of 120 scripts which were re-marked. The marks for General Impression may be considered for certain purposes as the most important of all, because they were allotted after the examiners had considered all the other categories, and because the examiners agreed generally that marks for General Impression were the marks they would have awarded at a School Certificate examination had they been free from the obligation to pass or give credit to a stated percentage of the candidates (see para. 35A, *ad fin.*).

## STATISTICAL REPORT

71

TABLE XIX. (1)  
COMPARISON OF 1ST AND 2ND MARKINGS FOR GENERAL IMPRESSION  
ESSAY-SUBJECT A

Cand.	Examiner P		Examiner Q		Examiner R		Examiner S		Examiner T		Examiner U		Examiner V		Examiner W				
	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo	
	Marking		Marking		Marking		Marking		Marking		Marking		Marking		Marking		Marking		
5	8	7	-1	10	7	-3	10	8	-2	10	7	-3	0	8	+2	8	9	+1	
7	7	8	+1	10	8	+2	11	7	-4	9	8	-1	7	7	0	10	8	-2	
9	7	7	0	7	8	+1	10	6	-4	5	7	+2	6	6	-2	11	11	0	
11	6	9	+3	10	8	-2	7	6	-1	0	8	-1	7	10	-1	10	9	-2	
14	12	12	0	10	11	+1	11	13	+2	10	9	-1	7	10	-2	7	10	-2	
23	9	8	-1	10	10	+0	8	12	+4	10	19	+0	6	10	+2	10	10	+1	
25	7	7	0	8	9	+1	7	7	0	5	7	+2	5	9	+4	9	10	+2	
31	8	8	0	7	8	+1	8	7	-1	9	9	0	5	9	+4	7	7	0	
36	7	6	-1	6	7	+1	6	6	0	5	7	+2	5	7	-2	10	0	-1	
38	6	6	0	7	7	0	5	6	+1	6	8	+2	3	7	+2	7	7	0	
43	2	2	0	3	3	0	5	6	-3	2	5	+3	3	4	+4	9	8	-1	
49	6	6	0	8	6	-2	6	5	-1	5	7	+2	6	7	+1	6	7	+1	
67	10	9	-1	9	9	0	8	8	0	9	11	+2	8	0	-2	7	7	+2	
61	10	10	0	11	10	-1	11	12	+1	12	12	0	11	12	-1	11	8	-1	
64	6	8	+2	8	8	0	7	8	+1	10	11	+1	6	9	+4	8	8	0	
72	7	8	+1	7	8	+1	6	6	0	6	8	+2	5	7	-1	8	7	+2	
81	12	12	0	12	13	+1	12	11	-1	8	9	+1	10	11	+1	12	0	-1	
89	7	7	0	8	9	+1	8	8	0	9	8	-1	6	10	+4	8	9	+2	
93	6	6	0	9	8	-1	6	7	+1	9	9	8	-1	11	12	+1	10	0	-1
97	12	12	0	9	11	+2	12	12	0	12	11	-1	11	13	+2	9	11	+1	
Average difice, neg- lecting signs		0.55	1.05		1.35		1.40		2.10		2.15		1.00		1.16		1.16		

## THE MARKING OF ENGLISH ESSAYS

TABLE XIX (2)  
COMPARISON OF 1ST AND 2ND MARKINGS FOR GENERAL IMPRESSION  
ESSAY-SUMMARY A1

Cand.	Examiner P		Examiner Q		Examiner R		Examiner S		Examiner T		Examiner U		Examiner V		Examiner W			
	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo
	Marking				Marking				Marking				Marking				Marking	
1	6	6	0	7	6	-1	6	4	-2	8	8	+1	6	9	+3	7	8	+1
6	8	8	0	9	9	+0	6	7	+2	6	9	+1	10	7	-3	-1	11	-1
11	4	7	+3	8	9	+1	6	6	-1	6	7	+1	6	11	+5	-1	8	+3
13	6	6	+1	7	8	+1	6	5	-1	7	9	+2	7	8	+1	7	0	+3
20	7	8	+1	8	7	-1	7	7	0	8	8	+1	6	7	+1	7	7	-1
28	6	7	+1	8	7	-1	6	6	0	5	5	0	7	11	+4	5	6	-4
35	9	9	+0	11	8	-3	11	8	-3	8	8	0	6	9	+3	8	7	-1
44	7	9	+2	7	9	+2	8	11	+3	5	6	+1	4	8	+4	7	8	+2
52	7	8	+1	7	7	0	8	5	-3	9	7	-1	6	7	+1	6	7	-1
54	8	7	-1	10	9	-1	11	10	-1	8	8	0	8	10	+2	9	11	+2
55	6	5	-1	9	6	-3	5	5	0	5	5	0	4	7	+3	7	8	-1
59	7	9	+2	9	8	-1	9	7	-2	6	8	+2	4	10	+6	10	9	-1
60	6	7	+1	8	9	+1	8	6	-3	6	7	0	7	10	+3	9	10	0
61	10	12	+2	11	11	0	12	11	-1	11	10	-1	10	13	+3	10	11	+2
69	2	4	+2	5	5	0	4	2	-2	2	7	+5	3	7	+4	6	7	-1
72	3	7	+4	9	7	-2	8	8	0	6	6	0	3	7	+4	8	6	-4
74	8	10	+2	8	9	+1	7	6	-1	5	7	+2	9	11	+2	9	7	+3
83	9	7	-2	9	8	-1	9	6	-3	6	7	+1	7	9	+2	8	6	-2
85	11	8	-3	13	8	-5	12	10	-2	6	7	+1	5	8	+3	10	9	+1
86	10	10	0	8	8	0	7	7	0	6	7	+2	5	10	+1	11	13	+2
Average diffo, neg.	lecting signs		1.45	1.25		1.45	1.10		3.05	1.45		1.10	1.10		1.10	1.00		1.00

TABLE XIX (3)  
COMPARISON OF 1ST AND 2ND MARKINGS FOR GENERAL IMPRESSION  
ESSAY-SUBJECT B

## THE MARKING OF ENGLISH ESSAYS

TABLE XIX (4)  
COMPARISON OF 1ST AND 2ND MARKINGS FOR GENERAL IMPRESSION  
ESSAY SUBJECT B1

TABLE XIX (6)  
COMPARISON OF 1ST AND 2ND MARKINGS FOR GENERAL IMPRESSION  
ESSAY-SUBJECT C

Cand.	Examiner P		Examiner Q		Examiner R		Examiner S		Examiner T		Examiner U		Examiner V		Examiner W			
	1st	2nd	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo	
	Marking	Marking	Marking															
2	7	8	+1	7	8	+1	5	6	+1	8	8	0	7	9	+2	0	7	+3
6	8	8	+0	10	8	-2	7	7	0	11	10	-1	9	8	-2	8	9	+1
30	9	10	+1	10	8	-2	8	8	-0	8	8	0	7	9	-1	0	9	-1
31	11	8	-3	10	10	0	9	8	-1	8	8	0	11	10	-1	7	10	+3
32	6	7	+1	5	6	+1	7	7	0	8	8	0	11	10	-1	7	8	+1
35	8	8	0	7	8	+1	6	8	0	8	10	+2	5	7	+2	7	4	-3
41	11	12	+1	12	12	0	10	8	-2	10	10	0	10	12	+2	11	7	-6
45	7	7	0	10	10	-0	8	12	+4	9	9	0	8	11	+3	9	11	+2
50	8	7	-1	9	8	-1	8	6	-2	8	8	0	9	9	-2	9	11	-1
62	7	7	0	9	10	+1	7	7	0	10	8	-2	6	9	+3	10	9	-1
65	5	6	+1	8	7	-1	5	5	0	8	8	-0	6	9	+3	10	9	-1
66	7	8	+1	10	11	+1	8	8	0	8	8	0	7	9	+3	9	7	-3
63	7	7	0	10	10	0	8	8	0	11	10	-1	8	10	+2	9	9	0
69	5	6	+1	10	8	-2	5	6	+1	7	7	0	6	8	+2	10	8	-2
76	8	8	0	10	8	-2	7	8	+1	8	9	+1	9	9	0	10	8	-1
78	7	8	+1	10	10	0	7	7	0	9	9	0	10	9	-1	8	8	+1
90	7	10	+3	11	8	-3	8	6	-2	10	10	0	10	8	-2	9	9	0
93	5	7	+2	8	6	-2	6	5	-1	7	7	0	6	7	+2	7	9	+2
94	8	10	+2	9	8	-1	8	8	0	12	11	-1	10	9	-1	7	7	-1
95	11	11	0	11	11	0	9	8	-1	8	10	+2	11	12	+1	11	10	-1
Average diffo, neglecting signs		*	0.95		1.05		0.80		0.50		1.75		1.60		1.15		1.05	

## THE MARKING OF ENGLISH ESSAYS

TABLE XIX. (6)  
COMPARISON OF 1ST AND 2ND MARKINGS FOR GENERAL IMPRESSION  
ESSAY-SUBJECT C1

Cand.	Examiner P			Examiner Q			Examiner R			Examiner S			Examiner T			Examiner U			Examiner V					
	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo	1st	2nd	Diffo			
16	8	10	+2	12	12	0	11	8	-3	8	8	0	10	10	0	11	11	0	10	11	+1			
18	7	8	+1	8	0	+1	6	7	+1	7	11	+4	8	9	+2	8	8	0	8	8	0			
25	11	0	-11	10	9	-1	8	-1	9	8	-1	10	13	+3	0	8	9	+1	8	9	+1			
29	10	10	0	10	10	0	10	7	-3	12	10	-2	11	11	0	11	11	0	11	11	0			
33	8	8	0	8	10	+2	9	10	+1	11	10	-1	10	10	0	10	8	+2	9	8	-1			
37	8	7	-1	8	9	+1	8	7	-1	8	8	-1	6	6	0	10	10	0	8	7	-1			
45	8	7	-1	9	10	+1	11	7	-4	12	10	-2	9	10	+1	7	7	0	12	14	+2			
54	7	6	-1	10	11	+1	7	6	-1	11	7	-1	10	10	0	10	9	-1	7	7	0			
56	6	6	0	9	9	0	8	9	+1	11	10	-1	10	9	-1	9	9	0	8	8	-1			
59	7	6	-1	9	9	0	7	6	-1	8	8	-1	6	7	+1	9	9	0	11	11	0			
66	8	9	+1	7	8	+1	7	6	-1	7	8	+1	6	7	-1	8	7	-1	12	11	-1			
68	8	9	+1	13	14	-1	11	11	0	12	11	-1	12	14	+2	12	11	-1	12	14	+2			
70	8	7	-1	9	8	-1	11	9	-1	8	9	+1	7	8	+1	8	8	0	11	11	0			
75	11	11	0	11	10	-1	11	8	-3	10	11	+1	11	12	+1	11	11	0	12	11	-1			
76	7	7	0	8	8	0	8	7	-1	6	7	+1	6	7	-1	8	7	-1	7	7	0			
82	8	7	-1	8	9	+1	0	6	-3	6	7	+1	6	8	+2	8	9	+1	7	7	0			
83	11	7	-4	12	11	-1	9	0	-1	9	7	-2	8	10	+2	10	12	+2	9	9	0			
90	8	6	-2	11	10	-1	0	9	0	7	8	-2	5	7	+2	10	12	-1	6	6	0			
95	11	7	-4	11	11	0	0	8	0	10	8	-2	7	12	+5	11	11	0	-1	0	10			
100	0	8	-1	11	11	0	8	0	+1	10	10	0	10	10	-1	8	10	+2	10	9	-1			
Average diffo, neg. lecting signs																								
	1.20	0.70	1.45				0.55	2.30		0.55			1.10			0.55			1.10			0.75		

103A. In Table XIX (7) are summarised the number and magnitude of the 120 differences between the first and second markings of the different examiners for General Impression.

TABLE XIX (7)

NUMBER AND MAGNITUDE OF DIFFERENCES BETWEEN THE FIRST AND SECOND MARKINGS OF THE DIFFERENT EXAMINERS FOR GENERAL IMPRESSION

Differences between 1st and 2nd Marking	Examiners							
	P	Q	R	S	T	U	V	W
0	51	34	35	42	9	34	26	31
1 unit	44	57	48	43	28	48	50	52
2 units	15	19	16	30	33	29	27	22
3 units	7	8	14	4	24	7	10	9
4 units	3	1	7	0	21	2	6	5
5 units	0	1	0	1	4	0	1	1
6 units	0	0	0	0	1	0	0	0
	120	120	120	120	120	120	120	120

Two of the examiners differ from the rest, Examiner P in his consistency and Examiner T in his want of consistency. The number of cases out of 120 in which the difference between the two markings of an examiner does not exceed one unit on the fifteen-point scale (6.67 on a hundred-point scale) is shown below :—

P	Q	R	S	T	U	V	W
95	91	83	85	37	82	76	83

It is to be remembered that a unit difference may be positive or negative and that within the group for which an examiner's differences are of one unit only, two candidates who were marked by him as equal in the first instance may be awarded marks differing by two units (or over 13% on a 100-point scale) in the second instance. (Take for instance P's marks for candidates No. 36 and No. 72 for Essay A.) The second markings even of the most consistent Examiner P differ in three instances by four units (or 26.6% on the hundred-point scale) from his first markings. Differences such as these may very seriously affect the fate of candidates.

103B. The following tables show the average differences between the first and second markings of 120 essays for the category General Impression. In Table XIX (8) the signs are included in calculating the averages ; in Table XIX (9) they are all taken as positive.

TABLE XIX (8)

AVERAGES OF THE DIFFERENCES BETWEEN THE 1ST AND 2ND MARKINGS  
OF 8 EXAMINERS FOR GENERAL IMPRESSION (20 SCRIPTS FOR EACH  
ESSAY-SUBJECT)

[*The signs of the deviations from the average are included*<sup>1</sup>]

Examiner	P	Q	R	S	T	U	V	W	Average
Essay A	+0.15	-0.05	-0.35	+0.50	+2.10	-0.40	+0.45	-0.35	0.26
" A1	+0.75	-0.65	-0.95	+0.80	+3.05	-0.25	+0.80	-0.30	0.41
" B	+0.40	-0.65	-1.05	+0.75	+2.50	+0.50	+1.50	+0.90	0.61
" B1	+0.20	+0.40	-1.00	-0.60	+1.80	+0.20	+0.35	-0.05	0.16
" C	+0.55	-0.55	-0.10	0.00	+1.45	-1.50	+1.05	+0.05	0.12
" C1	-0.70	+0.10	-1.05	-0.05	+2.10	-0.30	+0.70	+0.15	0.12
Average	+0.23	-0.23	-0.75	+0.23	+2.17	-0.29	+0.81	+0.07	0.278

<sup>1</sup> The + sign indicates that the second marking was greater than the first.

TABLE XIX (9)

[*All the signs of the deviations from the average have been taken as +*]

Examiner	P	Q	R	S	T	U	V	W	Average
Essay A	0.55	1.05	1.35	1.40	2.10	1.00	1.25	1.15	1.23
" A1	1.45	1.25	1.45	1.10	3.05	1.45	1.10	1.90	1.59
" B	0.90	1.35	1.25	0.95	2.50	1.20	1.90	1.50	1.44
" B1	0.30	1.00	1.20	1.10	2.10	1.00	1.65	1.05	1.18
" C	0.95	1.05	0.80	0.50	1.75	1.60	1.15	1.05	1.11
" C1	1.20	0.70	1.45	0.95	2.30	0.50	1.10	0.75	1.12
	0.89	1.07	1.25	1.00	2.30	1.13	1.36	1.23	1.278

#### Remarks :

(i) Table XIX (8) shows that Examiners T and V gave higher marks on the second marking for all six essays. Examiner R, on the other hand, gave lower marks throughout. In each of these cases the average difference is much greater than the average difference of all the examiners (0.278). The variability of Examiners T, V and R is also shown in Table XIX (9), which gives the absolute values of the differences regardless of sign.

(ii) On the whole, the examiners have tended to mark Essays A1 and B much higher on the second markings.

(iii) Omitting from the above tables the marks of Examiner T, whose awards on the second reading were more than 2 marks out of a maximum of 15 higher than on the first, the differences for the remaining seven examiners are:-

	Average Differences with Sign	Average Differences with all Signs taken as positive
Essay A	-0.007	1.11
" A1	+0.029	1.39
" B	+0.336	1.29
" B1	-0.071	1.04
" C	-0.071	1.01
" C1	-0.164	0.95
Average	0.018	1.132

103c. The average ranges for General Impression (i.e. the averages of the differences between the highest and the lowest marks for that category of the different examiners for the different scripts) were as follows :—

Average Ranges at the First Marking of 600 scripts. Essay-subjects : A, 4.03 ; A1, 4.44 ; B, 3.79 ; B1, 3.81 ; C, 3.74 ; C1, 3.63.

General average for the 600 essays (on the fifteen point scale), 3.91.

Average Ranges at the Re-marking of 120 scripts out of the 600. Essay-subjects : A, 3.15 ; A1, 4.10 ; B, 4.30 ; B1, 4.45 ; C, 3.55 ; C1, 3.90.

General average for 120 essays (on the fifteen-point scale), 3.91.

---

The exactness of the agreement of the general averages in the two cases is of course due to chance.

Examples of the actual marks awarded by the different examiners will be found in Tables II (1) to II (6); in Tables XVII (1) to XVII (6); and, for 120 essays, both at the First Markings and, at the Re-marking, in Tables XIX (1) to XIX (6).

104. The marks for Literacy have been considered separately (see pp. 80-92 below). It will be seen from Table XXV on p. 85 below that out of every six verdicts one was changed (159 out of 960). The largest number of changes was from "Doubtful" in the first marking to "Yes" in the second marking—90 in all. There were 21 changes from "No" to "Yes," and 5 changes from "Yes" to "No." Examiner T made far more changes than any other examiner.

105. *Conclusions* : We may summarise as follows the conclusions derived from the results of re-marking :—

(1) The consistency of marking as measured by correlation coefficients varies with the essay-subject, but there is little evidence of differences in consistency of marking between the directed and the undirected essays.

(2) Examiners' marks awarded to the same script on two occasions vary considerably, in some instances, both in the standard deviation and the mean as well as in the rank order of the marks.

(3) The consistency of marking varies with the examiner and the results emphasise the need for choosing examiners carefully.<sup>1</sup>

<sup>1</sup> It should be added that, so far as we are aware, examining bodies at present do not regularly test the consistency of the examiners in the way in which they were tested during the present investigation.—P.J.H.

## CHAPTER V.—LITERACY AS A CRITERION AT THE SCHOOL CERTIFICATE STAGE

106. In an official report on the School Certificate Examination published in 1932 the following sentences occur :—

(1) The Investigators are in agreement with those Examining bodies who take steps of one kind or another . . . to prevent the award of certificates to candidates who can only be classed as "illiterate."

(2) In the case of candidates taking English as a group subject they would have no difficulty in detecting and sentencing sheer "illiteracy."<sup>1</sup>

The objective of the Investigators seems a reasonable one at first sight.<sup>2</sup> But, when we try to define the words "literate" and "illiterate," and "literacy" and "illiteracy" in their application to candidates for the School Certificate and their work we are confronted with difficulties not immediately apparent.

107. The Oxford English Dictionary assigns a number of meanings to the word "literate" and its opposite "illiterate," of which we need only discuss two.

The word "literate" may mean a "liberally educated or learned person"; and it may have the widely different and much more frequent meaning of a person who "can read and write —opposed to illiterate." Neither of these meanings would be

<sup>1</sup> The School Certificate Examination; being the Report of the Panel of Investigators appointed by the Secondary Schools Examinations Council to enquire into the Eight Approved School Certificate Examinations held in the summer of 1931 (H.M. Stationery Office, 1932), para. 70. Quoted in *Essays on Examinations*, by Sir Michael Sadler and others (Macmillan, 1936), p. 142.

<sup>2</sup> The particular methods which the Investigators recommended for attaining their object were not adopted. In Circular 1463 of the Board of Education of 18 July, 1938, it was recommended that success in an English language test (which had been previously required by the Northern Universities Joint Board) should be treated as a condition for the award of the School Certificate, and this requirement was adopted by seven of the School Certificate Authorities in 1939 and by the remaining one in 1940.

of any practical value in judging the work of candidates for the School Certificate Examination, for the simple reason that, broadly speaking, none of such candidates would be literate in the first sense, and all of them would be literate in the second.

108. The problem which we set ourselves was this :—

Is it possible to devise a definition of literacy suitable for the particular purpose of the Investigators ? As has been stated above, our Sub-Committee and our examiners found the task beyond their powers. They could not agree. Nevertheless it seemed possible that, although the examiners could not agree on a verbal definition of the requirements which they had in mind, they might, in practice, find themselves able to reach a fair degree of agreement in judging whether any particular script satisfied their literacy requirements or not. If they could find such a reasonable measure of agreement in deciding whether any particular School Certificate essay did or did not show evidence of literacy, the absence of a definition, though regrettable, would not be fatal to the use of the word literacy for the purpose desired by the Investigators. The matter was therefore put to the test, and each one of our 600 essays was marked independently for literacy by each of the eight examiners. At a later stage, 120 of these essays were re-marked for literacy, as well as for the other seven categories, by the same examiners ; see paras. 117-122 below.

109. The examiners felt that it would be useless to adopt for literacy the fifteen-point scale used in marking the other categories, and for this further category each essay was therefore marked on a three-point scale as "Yes," "?" (i.e., "Doubtful"), or "No."

110. When the marks thus awarded to the scripts by the eight examiners were either

- (i) Yes, Yes, Yes, No, No, No, ?, ?, ?
- or (ii) Yes, Yes, Yes, No, No, ?, ?, ?, ?
- or (iii) Yes, Yes, No, No, No, ?, ?, ?, ?

these scripts were considered as showing a maximum of only three examiners in agreement ; this constitutes, of course, the widest possible *disagreement* between the eight examiners using the system above described.

The number of cases for which there were 3, 4, 5, 6, 7 and 8 examiners in agreement are shown in Table XX below.

TABLE XX

NUMBERS OF CASES IN WHICH DIFFERENT NUMBERS OF EXAMINERS  
AGREE IN MARKING FOR LITERACY

Examiners who agree See para. 74	Maximum number of Scripts for each Essay-subject on which the Examiners agree					
	A	A1	B	B1	C	C1
3	4	3	3	1	2	2
4	8	9	12	6	5	7
5	15	17	13	13	9	9
6	15	15	13	21	10	12
7	29	27	15	13	11	16
8	29	29	44	46	63	54
	—	—	—	—	—	—
	100	100	100	100	100	100
	—	—	—	—	—	—

It will be seen that the judgment of the examiners in regard to the literacy of the candidates varied with the subject of the essay. The number of cases in which the whole eight examiners agreed in their awards was only 29 for subjects A and A1, but rose to 54 for subject C1 and to 63 for subject C.<sup>1</sup>

111. We have not attempted to ascertain in any detail whether the numerical results showed any sex-difference in respect of literacy, but the following statistics of candidates who received the award of "Yes" for literacy from all the examiners indicate a superiority of the girls in this respect.

TABLE XXI

NUMBER OF CANDIDATES MARKED "YES" FOR LITERACY BY ALL THE EIGHT EXAMINERS

Essay-subject	Girls	Boys
A	18	11
A1	18	11
B	27	17
B1	26	20
C	38	25
C1	31	23
	—	—
	158	107
	—	—

112. Table XXII below shows the percentages of the three literacy awards "Yes," "No," "Doubtful," allotted by the examiners for the different essay-subjects.

<sup>1</sup> These scripts were not examined in the order A, A1, B, B1, C, C1; so that the increasing agreement from A to C was not due to increasing practice.

LITERACY AS CRITERION AT SCHOOL CERTIFICATE STAGE 43

TABLE XXII

PERCENTAGES OF THE THREE AWARDS FOR LITERACY ALLOWED BY THE  
DIFFERENT EXAMINERS FOR EACH ESSAY-SUBJECT

Essay-subject	Mark	Examiners							
		P	Q	R	S	T	U	V	W
A	Yes	70	90	68	87	34	88	96	61
	Doubtful	21	4	11	4	46	6	3	5
	No	9	6	1	9	20	7	1	3
A1	Yes	73	87	92	82	39	59	63	31
	Doubtful	17	9	6	11	42	18	11	77
	No	10	4	2	7	10	23	6	12
B	Yes	68	91	90	85	52	86	92	74
	Doubtful	25	8	9	12	39	8	6	12
	No	7	1	1	3	9	0	2	11
B1	Yes	78	93	94	89	49	79	95	78
	Doubtful	18	3	5	6	43	12	2	8
	No	4	4	1	5	8	9	0	14
C	Yes	73	95	91	93	72	97	99	81
	Doubtful	19	3	6	2	21	0	1	7
	No	8	2	3	5	7	3	0	12
C1	Yes	79	96	94	91	63	94	100	72
	Doubtful	16	2	5	3	20	1	0	12
	No	5	2	1	6	8	5	0	16

113. Table XXIII below shows the ranges of the percentages for each examiner and each award.<sup>1</sup>

TABLE XXIII

Err. P	Maximum	Minimum	Range	Err. Q	Maximum	Minimum	Range
Yes	79 (C1)	68 (B)	11	Yes	96 (C1)	87 (A1)	9
Doubtful	25 (B)	16 (C1)	9	Doubtful	9 (A1)	2 (C1)	7
No	10 (A1)	4 (B1)	6	No	6 (A)	1 (B)	5

Err. R	Yes	Err. S	Yes	Err. T
Yes	94 (B1) & C1) 88 (A)	Yes	93 (C)	82 (A1) 11
Doubtful	11 (A) 5 (B1) & C1)	Doubtful	12 (B)	2 (C) 10
No	3 (C) 1 (A, B, B1, C1) 2	No	9 (A)	3 (B) 6

<sup>1</sup> The symbol within parentheses following a percentage shows the essay-subject for which the percentage was given by the examiner in question.

TABLE XXIII—*continued*

<i>Exr. T</i>	<i>Maximum</i>	<i>Minimum</i>	<i>Range</i>	<i>Exr. U</i>	<i>Maximum</i>	<i>Minimum</i>	<i>Range</i>
Yes	72 (C)	34 (A)	28	Yes	97 (C)	59 (A1)	38
Doubtful	46 (A)	21 (C)	25	Doubtful	18 (A1)	0 (C)	18
No	20 (A)	7 (C)	13	No	23 (A1)	3 (C)	20
<i>Exr. V</i>				<i>Exr. W</i>			
Yes	100 (C1)	83 (A1)	17	Yes	81 (C)	61 (A)	20
Doubtful	11 (A1)	0 (C1)	11	Doubtful	12 (A1, B, C1)	7 (C)	5
No	6 (A1)	0 (B, B1, C1)	6	No	31 (A)	11 (A1)	20

114. Table XXIV below shows the wide differences of standard of the different examiners in the form of the percentages of awards of "Yes," "Doubtful," and "No" which they assigned to all the essays taken together (on their first marking).

TABLE XXIV

## Examiners' Percentages

	P	Q	R	S	T	U	V	W	Average percent- age
Yes	73.5	92.0	91.5	87.8	51.5	83.8	94.7	73.8	81.1
Doubtful	19.3	4.8	7.0	6.3	36.7	7.3	3.8	9.8	11.9
No	7.2	3.2	1.5	5.8	11.8	8.8	1.5	16.3	7.0

Taking the two lowest and the two highest percentages for each award and for the different examiners, we find the following :—

	<i>Lowest Percentages</i>	<i>Highest Percentages</i>
"Yes"	T, 51.5%; W, 73.8%	Q, 92.0%; V, 94.7%
"Doubtful"	V, 3.8%; Q, 4.8%	P, 19.3%; T, 36.7%
"No"	R, 1.5%; V, 1.5%	U, 8.8%; T, 11.8%

It will be obvious from the foregoing tables that for Examiners Q, R, and V, the word "literate" has a widely different connotation from that which it has for Examiners P and W, and that these again differ no less widely from Examiner T.

On the whole number of awards, the percentage of "Noes" was 7; of "Doubtfuls," 11.9, and of "Yeses," 81.1.

115. Thus, taking the average, nearly 19 per cent. of the verdicts of the examiners were "No" or "Doubtful." In so far as any significance can be attached to results so discrepant for the different examiners, the figure supports the severity of the judgment of the Government Investigators on the standard of English displayed at the School Certificate examination.

To ensure greater efficiency the methods used at the examination have been recently changed at the suggestion of the Board

of Education, though not according to the method recommended by their Investigators (see para. 106, footnote 2, above).

116. Only one certain conclusion seems possible from the foregoing figures, namely, that at the School Certificate stage the term "literacy" has no meaning sufficiently clear and well-recognised for it to be used effectively in judging the scripts of the candidates.

### *Re-marking for Literacy*

117. It will be remembered (see paras. 36, 44 and 96-105 above) that after an interval of five months each examiner re-marked 120 essays out of the 600 (the same for all examiners). The general results of the re-marking have been analysed separately but it seems more convenient to deal, in this place, with the comparatively simple question of the re-marking for Literacy, which involves no complicated mathematical analysis.

118. Table XXV below shows the number of changes between the first and second markings and their nature.

TABLE XXV  
CHANGES BETWEEN FIRST AND SECOND MARKINGS FOR LITERACY

	Marking		Examiners																	
			1st	2nd	P	Q	R	S	T	U	V	W	P	Q	R	S	T	U	V	W
Essay A																				
Agreement	Yes	Yes	11	17	17	16	5	16	19	13			13	18	16	16	3	11	19	13
	?	?	4	-	-	-	5	1	-	-			-	-	-	1	-	-	2	
	No	No	1	1	-	1	1	-	-	3			1	-	-	-	-	-	1	
Change of 1 unit	Yes	?	1	1	-	-	1	-	-	-			-	1	3	-	-	-	-	1
	?	Yes	2	1	2	2	7	-	-	1			1	-	1	2	10	2	1	
	No	?	-	-	1	-	1	1	-	1			2	1	-	1	4	3	-	
	?	No	-	-	-	-	-	-	-	-			2	-	-	-	-	1	-	
Change of 2 units	No	Yes	1	-	-	1	1	-	1	2			1	-	-	-	3	2	-	1
	Yes	No	-	-	-	-	1	-	-	-			-	-	-	-	1	-	-	
Essay B																				
Agreement	Yes	Yes	11	17	19	17	7	13	17	15			12	19	20	19	11	16	20	17
	?	?	4	-	-	1	2	-	-	-			1	-	-	-	-	-	-	
	No	No	1	1	-	-	-	1	-	2			-	-	-	1	-	-	2	
Change of 1 unit	Yes	?	1	-	-	-	-	-	-	-			3	-	-	-	-	-	-	
	?	Yes	2	2	1	2	9	6	2	2			2	-	-	-	8	1	-	
	No	?	-	-	-	-	1	-	-	1			-	-	-	1	-	-		
Change of 2 units	No	Yes	-	-	-	-	1	-	1	-			-	1	-	-	3	-	-	
	Yes	No	1	-	-	-	-	-	-	-			-	-	-	-	-	-		

TABLE XXV—*continued*

Marking	Examiners										P	Q	R	S	T	U	V	W				
	1st	2nd	P	Q	R	S	T	U	V	W												
			Essay C										Essay Cl									
Agreement	Yes	Yes	15	20	17	18	16	18	20	18	16	20	20	20	12	20	20	16				
	?	?	—	—	—	1	—	—	—	—	2	—	—	—	1	—	—	—				
	No	No	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
Change of 1 unit	Yes	?	1	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—				
	?	Yes	1	—	3	—	4	—	—	—	1	—	—	—	7	—	—	3				
	No	?	2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1				
	?	No	—	—	—	1	—	—	—	1	—	—	—	—	—	—	—	—				
Change of 2 units	No	Yes	1	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—				
	Yes	No	—	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—				

119. Table XXVI below gives a summary of the changes, analysed according to essay-subjects.

TABLE XXVI  
SUMMARY OF NUMBER OF CHANGES IN RE-MARKING FOR LITERACY

Nature of change	Essay-subjects and number of changes						Total
	A	A1	B	B1	C	Cl	
Yes—Doubtful	3	5	1	3	1	1	14
Doubtful—Yes	15	18	26	12	8	11	90
No—Doubtful	4	12	2	1	2	1	22
Doubtful—No	0	3	0	2	2	0	7
Yes—No	1	1	1	0	2	0	5
No—Yes	6	7	2	4	2	0	21
Totals	—	—	—	—	—	—	—
	29	46	32	22	17	13	159

It will be seen that with 8 examiners and 120 essays there were altogether 960 verdicts on Literacy and that at the re-marking the examiners changed their minds in 159 cases, or, as nearly as possible, in one case out of six.

120. It will also be noticed that the number of changes varies considerably with the essay-subject—from 13 for essay-subject Cl to 46, more than three times the number, for essay-subject A1.

121. In Table XXVII below is shown a comparison of the percentages of the different marks for literacy allotted at the first and second markings of the different examiners.

TABLE XXVII

COMPARISON OF PERCENTAGES OF THE DIFFERENT AWARDS FOR LITERACY ALLOTTED AT THE FIRST AND SECOND MARKINGS BY THE DIFFERENT EXAMINERS

Percentage of "Yeses"	Examiner							
	P		Q		R		S	
	1st Mkg.	2nd Mkg.	1st Mkg.	2nd Mkg.	1st Mkg.	2nd Mkg.	1st Mkg.	2nd Mkg.
"Doubtfuls"	19.3	18.3	4.8	2.5	7.0	2.5	6.3	3.3
"Noes"	7.2	6.7	3.2	1.7	1.5	0.8	5.8	2.5
Examiner								
Percentage of "Yeses"	T		U		V		W	
	1st Mkg.	2nd Mkg.	1st Mkg.	2nd Mkg.	1st Mkg.	2nd Mkg.	1st Mkg.	2nd Mkg.
"Yeses"	51.5	86.7	83.8	90.0	94.7	100	73.8	86.7
"Doubtfuls"	36.7	12.5	7.3	5.0	3.8	0	9.8	5.8
"Noes"	11.8	0.8	8.8	5.0	1.5	0	18.3	7.5
							1st Mkg.	2nd Mkg.
							81.1	90.64

122. Every examiner raised the percentage of his "Yeses" at the second marking, though for Examiner P the change is not significant. Examiner V, the most lenient on the first marking, boldly declined to designate a single script as illiterate at the second.

The greatest change is in the figures of Examiner T who raised his percentage of "Yeses" from 51.5 to 86.7, and diminished the percentage of "Doubtfuls" from 36.7 to 12.5, and of "Noes" from 11.8 to 0.8. He had obviously no steady idea of what connotation should be attached to the term Literacy in marking School Certificate examination scripts.

*Typical examples of Essays regarded by the examiners as  
"illiterate" or of doubtful literacy*

123. Examples of some of the essays in regard to which there were marked differences in this matter between the examiners will show the difficulties with which they had to contend.

124. The following table shows the number of candidates who respectively received one, two, three, etc., "Noes" in the marking of each essay-subject.

TABLE XXVIII

Number of Candidates who received

Essay-subject	1 No	2 Noes	3 Noes	4 Noes	5 Noes	6 Noes	7 Noes	8 Noes	Total.
A	22	6	4	4	3	—	1	—	40
A1	20	10	3	—	—	3	1	1	38
B	15	4	3	—	1	1	—	—	24
B1	11	3	1	2	2	1	—	—	20
C	10	5	1	—	2	—	1	—	19
C1	11	4	4	—	1	1	—	—	21

The improvement from essay-subject A to essay-subjects B to C1, may be due in part to progress of the pupils during the course of the investigation. Table III, showing the mean marks for the essays in all categories, gives evidence of some improvement during the course of the investigation, but of nothing so marked as the foregoing table seems to indicate.

125. No pupil had all his essays marked "No" for literacy by all the examiners, but the weakest of all for literacy was No. 43. The following is a summary of his awards :—

Essay-subject	No. of "Yeses"	No. of "Doubtfuls"	No. of "Noes"
A	—	1	7
A1	—	—	8
B	—	2	6
B1	2	1	5
C	1	—	7
C1	2	—	6
Totals	5	4	39
	—	—	—

Out of the 48 verdicts given by the eight examiners on his six essays, there were 39 "Noes" and only 5 "Yeses." Samples of his work will therefore give some idea of what "illiteracy" means to School Certificate examiners who are in fairly close agreement in regard to particular scripts.

In the following passages the mistakes in the original are reproduced.<sup>1</sup> The essay on subject A1 by No. 43, the only one out of six hundred which all the eight examiners regarded as illiterate, is reproduced *in extenso* below.

#### No. 43.

##### GAMES IN SCHOOLS. A1

Games should be regarded as a pleasure not a duty, even in schools.

Games have always been a pleasure not another way to do work. Similarly games should not be regarded as making a living they also should not be taken too seriously.

The Australians are very good at games but they take it much too seriously. If they start loosing a game they accuse you of playing a dirty trick.

People may say that making games a duty improves discipline. When this point is reached games cease to be games they become a form of work. At this point quite a lot of pleasure is lost. If people try to force the pace of anything it cannot be so good as the work of a person who jogs along steadily. Similarly when discipline in a game

<sup>1</sup> It is to be noted that some mistakes, e.g., "lot" for "lost", are due to mere carelessness and not to ignorance, but no corrections of such mistakes have been made in the text.

is tightond up. interest is lot. Compare for example a Test match at cricket and a school match. In a Test match dicipline is not actually tighter but the poplo just treat it as a game. It should be like this in a school match somotimes in a school match you find a bullying type of captain who wants to be top dog overytime I do not say this happens everytimo but this typo of things spoils games.

Since a boy will never start at the top of a ladder he will have to take orders so he might just as well start on the playing field. Because when he starts out in the world. it will come as a great shock to the boy to take orders. Ho has loarnt to take orders in school a little but not to the extent that he will have to later on.

Games improvo the mind beyond doubt it makes you think. If you are not careful you will mako a boy hate games and they will become a bore to him this will happen I think if games are made a duty.

It will not improve a boys health if this happens. A person will engago a boy only if he has good health I think this comes before the point of having a school cirtificate.

Beyond doubt games should be regarded as a pleasure.

126. We quote the first thirteen and the last nine lines of 43's essay on subject C, which seven out of the eight examiners regarded as illiterate :—

#### SPEECH DAY

This is tho most memorial and auspicious occation of the year. It is a rendezvous for the boys at school, tho old boys who havo met tho difficulties of tho world and tho boys leaving school to go and fend for themselves. It is tho one occasion when the Head-master speaks to the boys when various goveners and tho boys parents are present among the assembly.

The evening opens with the song or hymn Jerusalem after which the chair man adresses the assembly. Then comes the Headmasters address to which every parent looks forward to. ho speaks of the comming events of any specially outstanding persons of the school and of any school viotories in the athletic side.

It is also bad for the health sitting in a crowded hall for some of the people havo to stand. The boys should have the choice of going to the speech day if they wanted to or not going. If this was tried over half the school would stop away and those that did go would chiefly be the prise-winners and those who wanted to see what kind of an ordeal it was.

Speech day should only be for the parents and the boys who wished to go.

127. Here are two passages from No. 43's essay on subject C1 (for which he had only six " Noes " and two " Yeses ").

As I entered the hall I was struck with the dreary ness of it since it was largo and gloomy. Around the walls were the remains of previous festvities and dances. These were in the form of paper-chains and coloured papers. In the corners were parts of the armie's equipment

which was uncurtained hid beneath large tarpaulins. The platform upon which visitors and important people stood was gaily coloured while a cover of satin and velvet hid the stains and marks on the old army table underneath. Row upon row of empty chairs were also seen in the morning but in the evening all these were full of murmuring people.

This year speech day was more interesting than usual since a new school is being built also a swimming bath. There was much laughter over the swimming cup which the Headmaster gave to the school 7 years ago. The school has never won the cup although they have been very near to it. Beckingham who has won the cup has a bath. Perhaps it [*a slip for "if"*] we have a bath we shall win it.

128. Finally, we quote the beginning and the end of No. 43's essay on subject B1, for which he had five "Noes," two "Yeses," and one "Doubtful."

Since we are Englishmen we are proud of our country, and we think that England is by far the best country in the world in which to live. I agree with this, in the same way as you agree with people who say this about your country.

The governments that our people prefer are slightly different in their views, although both our Fatherlands want peace at any price.<sup>1</sup> With our Democratic government King and Queen everyone is happy. There is prosperity in our country, although we owe part of this to the manufacture of armaments.

England has bred many famous men and some of these men have invented many things. Although some people doubt whether it is good for mankind. The famous men are Scott Shakespere Masefield Bacon Haig Nelson Raleigh Drake not to mention Pope and George Elliot.

From this account my beloved French friend, I think you will appreciate England a little more than you did before. There are one or two things in these few pages that you cannot boast of, especially in the way and the nature of the British people they are allround sportsmen but they are never contented because when you become perfectly contented you and a clam are first cousins.

I remain,  
Your friend,  
B.B.

129. It is quite clear that No. 43, though the most illiterate of our candidates, has quite a respectable power of saying what he wants to say without any kind of ambiguity. His work affords an illustration of the statement in para. 107 that no School Certificate candidate is "unable to read and write." Indeed no pupil unable to read and write would be admitted to a School

<sup>1</sup> This was written in 1937.

Certificate class. It would seem that it was only by a stretch of the imagination or of their own use of English that the Investigators could describe some School Certificate candidates as "illiterate."<sup>1</sup>

But the fact that our own examiners regarded a number of the essays before them as illiterate affords some justification of the exaggeration of the Investigators.

130. It may be interesting to quote passages from essays in which the examiners were far more divided in their judgments upon literacy than in the case of No. 43.

Pupils capable of writing some paragraphs of fair merit may flounder badly in others. It is not easy to find any single adjective comparable with either "literate" or "illiterate," adequate to describe such unequal work. Perhaps the word that we have chosen—"doubtful"—describes it most suitably; "uncertain" might also have its advantages in describing the performance of a person who scores two or three bull's eyes, one or two "outers," and actually misses the target altogether at times.

131. For his essay on subject A, No. 49 received three "Yeses," three "Doubtfuls," and two "Noes." We quote three passages marked (a), (b), (c), which illustrate the difficulties which the essay presented to the examiners in arriving at a verdict:—

(a) School life would be very dull if no time was given up to play games. Some schools have more time for games than others, and in the end these schools make better sportsmen of the boys and this is a great thing in a boy's life.

Some schools have the games period in the morning and some in the afternoon, and I think the morning games are the best, because it is cooler and the air is much more refreshing.

In most schools there are organised games with other schools, but only in certain games; every school has some kind of a gymnasium, in here not only do exercises take place for improving the body but also

<sup>1</sup> The average marks awarded to No. 43 for Sense and General Impression were as follows:—

Essay-subject	A	A1	B	B1	C	C1
Sense	4.1	4.0	4.75	6.1	4.0	5.4
General Impression	3.4	3.4	4.5	4.5	3.6	5.0

It is noteworthy that his average marks for Sense are consistently higher than those for General Impression. There is a good deal of variety in the individual judgments. Out of the 48 marks for Sense he gets seven 2's and five 7's and one 8, and out of the 48 marks for General Impression he gets ten 2's and only one 7. It may certainly be asked whether his marks for Sense were not affected by his deficiencies in other categories, although the script was marked for Sense before the other categories.

<sup>2</sup> According to Miss Vera Brittain, one Oxford examiner described the examination papers of that distinguished young writer, Winifred Holtby, as "quite illiterate" (*The Testament of Friendship*, 1940, p. 109). There is obviously a "sliding scale" in the current use of the word.

games, such as hand ball. In these organised games it would be much more interesting if there was some kind of a league formed.

(b) All kinds of sports should be encouraged, even if the school does not permit it, for instance some schools play rugby, and only rugby can be played on the sports field, why not put a soccer pitch in place of one of the rugby ones.

(c) These games should be encouraged much more than are ; not only would it make the boys good sportsmen, but also make them stronger and fitter than are today.

132. No. 55, for her essay on subject C1, was awarded four "Yeses," three "Noes," and one "Doubtful." We quote from it two passages marked (a) and (b) :—

(a) We were all in our seats long before all the important people were seated on the platform. The platform really looked lovely, it being decorated with ferns and flowers. When at last people did appear on the platform, the noise which had been going on in the hall ceased instantly. It was rather a loud noise and the parents said afterwards that it was really surprising that such a noise could suddenly be subdued to silence in a second.

(b) The Headmistress's report came next through which we had to do the usual amount of clapping. I am afraid that I cannot write down all what she said as it would take to long.

Next various other speeches were made by different people on the platform, all of which were loudly clapped.

The actual giving of the prizes came next and the long line of girls slowly received their prizes, this was so as each girl was loudly clapped, especially tho girls who had more than one prize.

Passage (a) explains the "Yeses," passage (b) the "Noes."

## CHAPTER VI.—EXAMPLES OF ESSAYS SELECTED ON THE BASIS OF THE MARKS ALLOTTED TO THEM BY THE EXAMINERS<sup>1</sup>

### (A) EXAMPLES OF THE BEST ESSAYS

133. In the chapter on Literacy we have given examples of essays regarded by the examiners as the poorest. We now quote examples of the work which they regarded as the best in the categories of Sense and General Impression. The examples have been taken mainly from those in the "representative selection" made by our statisticians for Tables II (1) to II (6), pp. 24–29 above. The marks for Sense and General Impression will precede each essay which we quote, in full or in part. The marks for the other categories are given in the tables referred to above.

134. For the undirected subject A (Games) we quote two essays, Nos. 61 and 81, which (by pure chance) express very different opinions. Both are by girls.

The marks are as follows :—

	Examiner	P	Q	R	S	T	U	V	W	Average
No. 61	Sense	12	11	12	13	12	12	12	12	12
	General Impression	10	11	11	12	12	12	12	11	11·4
No. 81	Sense	11	12	13	7	12	10	10	13	11
	General Impression	12	12	12	8	12	10	10	12	11

The awards of the different examiners are closely concordant for the two essays ; and but for the low marks given for the second essay by Examiner S, the averages for the two essays would have been, roughly speaking, only half a unit apart.

We print both essays *in extenso*.

No. 61.

#### ESSAY A

Games, to the average school girl or boy, form a recreation from the ordinary school work. This is not always so when the child has just begun school, because if asked which subject he prefers he will generally

<sup>1</sup> As in previous examples, mistakes in the originals have been reproduced. Proper names which were, or appeared to be, the names of real persons have been replaced by fictitious names or initials, in some cases enclosed in brackets. Addresses at the head of the letters have been treated in the same way.

choose the physical training period. In the case of small children this consists of simple singing games such as "Ring a ring of roses." However as the child grows he begins to be especially interested in a certain academic subject and no longer looks upon games as the most interesting lesson, although he has now graduated to hockey or football on the games field.

All children, unless they are ill, enjoy the games period ; but it is often difficult for the school-master to know how much "games" should be allowed. A school-master of one big school in England made the experiment of giving half his school twice as many games lessons as the other half and he found that too many games were conducive to bad examination results.

How then, should the games periods be allotted ? Should one whole afternoon be given each week or should physical training lessons be scattered over the whole week ? I think that the latter course is the most sensible. There is no greater incentive to work than the prospect of spending forty happy minutes on the games field after the more tedious lesson has finished. Therefore perhaps games always ought to finish the afternoon's work. However, consider the case of a boy or girl in a school team which has to practice after ordinary school hours. If the regular games lesson is the last lesson of the afternoon, they go out to practice with the team feeling tired.

The obvious answer is to have the games lesson in the last period of the morning ; which gives an answer to all our problems. It provides the incentive to work and then sends home the boy or girl ready for their lunch and a happy afternoon which is spent industriously because the pupil feels that she has had her enjoyment and will work hard for the rest of the day.

That is one question settled with regard to games, but another more important question arises. We have decided on the correct place for games in the syllabus of one day but how many times a week should this lesson occur ? How high should games rank in the training of young people ? In the example of the school master I have mentioned earlier it was found that too many games cause a falling off in other subjects. Many schools give three games lessons a week and this seems to be quite a satisfactory number, giving enough recreation to satisfy the pupils and yet not removing their interest too much from ordinary subjects.

There is no doubt that games does add to the mental ability ; for most games require alertness and quickness of thought which are also needed in other subjects. They give a young boy or girl training in self control and also foster a spirit of fair play which stands any person in good stead in after life. It is also well-known that a healthy body helps even the cleverest mind and games provides an opportunity for keeping fit. The competition which games provide is also good for young people and enables them to take defeat as well as success and to respect their opponents. Of course there are people who need food before they can do physical jerks and this has raised a storm of controversy for the Government are complaining of the physical fitness of the nation. In the poorer areas it is quite true that physical jerks can have no place until the children become well-nourished and happy, for this is the one necessary factor in the life of a child before it can really enjoy physical training and before physical training can take an important part in the daily school-life of every child.

No. 81.

## ESSAY A

If you told a little boy of about four years of age that instead of playing with his toy train he is to run races in the garden with his little sister, he would be puzzled and resentful. In the same way do some girls resent being forced to play games when they would rather be taking a quite walk or sitting under the trees in the school grounds, chatting. In spite of everything that is said in favour of compulsory organised games I will always maintain that they should not be compulsory. In the early years of school life they serve their purpose. They draw a form together and form a team spirit, and the younger girls enjoy them. I am speaking of rounders now when I say that having played this game for five successive summers I am utterly bored with it, and I think I am not alone when I express this opinion.

In the Upper IVth and Vth forms we have our games period in the last three quarters of an hour before we go home. In the summer this means that after sitting cooped up in school all day we are set free, as the staff fondly imagine, to enjoy ourselves. It is very hot, fielding in the sun and running, it is worse still. The sun beats down, we get hotter and stickier, and look longingly towards the cool, shady seats under the trees. No, we cannot go and sit down; we must keep up this farce of enjoying ourselves. We must pretend to be thrilled because someone has scored a rounder.

At the end of the game we are hot, tired and slightly irritable. Now, we have either to walk, cycle or ride in a stuffy bus, home. Are we then fit to settle down to two or three hour's homework? In the winter it is nearly as bad. We rush madly up and down a netball court and since it is cold we get a tight, choky feeling in our throats. It starts to rain a little, but we must keep on playing, and must not grumble because we want to be sports. What a farce it all is!

How far is this compulsory sport developing the so-called "Sporting Instinct"? Have you ever seen the crestfallen look of a not-very-good goal keeper when she has just let through a perfectly easy shot, nor the ill-concealed black looks of the rest of the team. They are not ill-natured girls—far from it, but they are disappointed. The goal-keeper should play in another place, you say? It is not so easy, because she just has to play where she is put, and cannot argue about it. "Someone has got to play goal keeper," the games-captain protests. It is a very cold place to play and therefore unpopular. Is that any reason why a girl who has not had much practice as goal-keeper should be subjected to that humiliation? Apparently "sport" covers a multitude of sins.

The old idea that all children should be made to learn the piano, however much they hate practising is dying out. Then isn't it time that compulsory games died a natural death? At present I am sitting in the formroom, writing this by the windows which are right down. It is almost like being out of doors, because it is so cool and [? all omitted by inadvertence] I can hear is the birds in the trees. I am much happier to be sitting here writing this essay than I would be if I were outside playing rounders or tennis. I am interested here. Outside I should be bored. . .

Of course, there are some girls in this form who love games. Let them play then. If we others prefer not to play why may we not sit out under the lovely trees and read or talk? We should be in the

open air and if three quarter's of an hour's rushing about is going to make the girls who are playing games into healthier women than I shall ever be, then I am willing to take back all have said and I will try to amass some enthusiasm for organised games ; but I am quite sure this will never be necessary. I am sure their games lesson will not profit them more than my quiet rest will profit me. I think they would slightly envy me and quite soon want to follow my example.

135. Subject A1, the "directed" essay on games in which pupils were asked to state pros and cons in the form of a speech to a Debating Society (see para. 18 above) was found much harder than subject A. The marks of No. 85 are as follows :—

<i>Examiner</i>	P	Q	R	S	T	U	V	W	Average
Sense	11	13	12	7	4	9	11	12	9.9
General Impression	11	13	12	.7	5	10	11	11	10

Both S and T mark the essay below the level of "competency," though the rest regard it as decidedly above it.

No. 85.

### ESSAY A1

#### SCHOOL GAMES

Mr. Chairman, ladies, the subject under debate today is whether games should be regarded as a duty or a pleasure. No doubt in the minds of most of you present here there is no question as to the answer to this question. However, there are usually two sides to every question so let us consider them carefully.

Most people nowadays enjoy games. There must be some reason for this and it is my opinion that it is the excitement of the game that gives them so much pleasure. It cannot be said with absolute truth that games are a source of relaxation from hard work. Games are harder on the physical strength of the player than an ordinary day's work and in some games an alert brain is essential for quick decisions which may alter the course of the whole game in an instant. Therefore games do make for a general alertness of the mind which is of considerable use in other spheres than that of sport. Games are not absolutely necessary for the development of the body but they are the easiest and by far the most pleasure method for children. Games are doubly pleasant after sitting doubled up at a desk for the most part of the day. It is a relief to stretch stiff muscles and to free in a fast game of tennis, cricket or hockey the energy which cannot be used up in study.

Games may also be regarded by some people as a duty more than a pleasure. They probably think that it is right for them to play games in order to carry out their daily work as well as possible. This does not mean, however, that because it is a duty they cannot obtain any pleasure from it. It is impossible for any healthy person to dislike games absolutely. You will notice that I said any healthy person. If the player is very delicate or cannot stand a fast game it is extremely unwise for him or her to try to do so. This would be carrying the idea of duty too far at the further detriment of the person's health. Other less violent ways of development should be sort by these unfortunate people to whom games mean nothing but misery.

Therefore, in conclusion, it is my opinion that games can be both a duty and a pleasure and should be regarded as such by those who wish to make their way in the world.

136. For the undirected subject B, "In Praise of England," the two pupils in the "representative selection" marked highest were Nos. 11 and 68, whose marks are as follows:—

	Examiner	P	Q	R	S	T	U	V	W	Average
No. 11	Senso	9	11	10	10	10	10	9	13	10.25
	General Impression	9	12	10	11	10	10	9	13	10.5
No. 68	Sense	11	10	11	9	12	12	11	12	11
	General Impression	11	12	11	9	12	12	11	12	11.2

As we quote other work of No. 68 we shall only quote here the essay of No. 11.

No. 11.

### ESSAY B

England is almost the only country in the world that at the present time has peace within its shores. In Europe Spain is divided in thought and is in the throes of Civil war—Italy and Germany are proud countries that have been cowed by their dictators who allow no freedom of thought or speech in their respective countries, while in England everyone is free and have not the fear of imprisonment hanging over their heads. France is fighting for its rights for the workers while Albania disapproves of their king. England's loyalty to its king was shown in last week's celebrations of the coronation. Our peace and goodwill was clearly visible in the affection everyone showed to their Majesties on Coronation Day.

England, although divided in opinions in the Government has a competent council of men to deal with the affairs of the country. The House of Commons understand the needs of the people because they are part of them. It is the Government of the People, by the people, for the people which makes England the peaceful country it is.

The people of England are loyal to their country which helps a lot for the ruling of this island. The British Army are faithful to their king and they protect and guard "this earth, this realm, this England." The navy also guard English shores and it is through them we have established a high position in the world. As well as being faithful to its king the inhabitants of England are faithful to God—as in the song of England "we highly dedicate, O Lord, to thee." Also that we "... may be renowned through all recorded ages for Christian service and true chivalry."

The countryside of England is very beautiful for almost every type of flora is found here. In South-East Kent rolling green downs and beautiful woods enhance the beauty of the high gleaming white cliffs and the treacherous blue-green sea that sparkles and flashes as the sun catches it. On the downs amidst the beech and ash trees one may hear in Spring, the chirruping of the birds and especially the beautiful song of the nightingale. Carried on a light breeze distant sounds

connected with a farm are often heard. The plaintive sound of a sheep admonishing its lamb for gambolling about—the anxious clucking of the hens and raucous noise of the cockerel—the low, deep bark of a sheepdog and the faint sound of the farmer taking the cows back to be milked. In the opposite part of England, the Northwest, one may see high, towering peaks—possibly with snow still on their summits—while down in the valley, nestled in a hollow, perhaps lies a quiet undisturbed town or lake perfect in its setting like a stone set in a ring—gleaming and glittering as ripples caused by a fish extend in an ever widening ring to the edge of the lake.

England of today is vastly different from that of yesterday, for it has obtained for itself a firm, strong position in the world which it had not got a hundred years ago and which I hope it will always possess.

137. For the "directed" essay on subject B1, Nos. 12 and 68 received the following marks:—

	Examiner	P	Q	R	S	T	U	V	W	Average
No. 12	Sense	8	13	11	14	14	10	11	12	11.5
	General									
No. 68	Impression	11	13	11	14	13	11	11	11	11.9
	Sense	11	11	13	11	12	12	12	13	11.9
	General									
	Impression	9	10	13	11	12	13	12	13	11.7

Both essays are printed *in extenso*.

No. 12.

### ESSAY B1

Dear Henry,

2nd July 1937

I was very interested indeed in your account of the annual village festivities which took place in your district a week ago. It gave me an insight into the social life of rural France which I had never had before. But I read with even greater pleasure what you said about the very conscious patriotism which is common to all French people, at all times of the year, whether at work or play; and I felt that I must, in turn, try to give you just a brief, and perhaps, rather vague, idea of that more silent and unobtrusive love-of-country which is peculiar to the Englishman, and which is, I believe, infinitely stronger than that of any other nationality.

In my opinion, the average Englishman is still, at heart, the old-fashioned John Bull—I think you must have heard of our now immortal John Bull—the lover of England's glorious past and her ever sweet-smelling countryside, of the yeoman and the newly ploughed acres, of the white-washed thatched cottage, with its roses and cabbage-plot, and—I must not forget them—of the good old roast beef and beer of Merrie Englande.

I do not want to argue over such material matters as, for instance, whether English or Continental cooking is the best, for such things depend wholly on individual taste. If I did, I expect you would soon point out that your roads are immeasurably finer than ours.

All I can do is to remind you of our great King, our democracy, our individual liberty, and our prosperity; these are some of our greatest possessions, which every true Englishman is most carefully preserving. I suppose it would be quite tactless to suggest to most French people that Napoleon's remark about a Nation of Shopkeepers was a high compliment, but it definitely was.

The English people are the most good-tempered on earth; intellectually we are tolerant; socially we are good-humoured. It is, I may say, this right good humour, illuminated as it is by reason, that constitutes one of the most characteristic and noble qualities of our race.

All over the globe we find Englishmen; mainly because in the home country we have the enormous advantage of being within a few miles of the sea—compared with other countries—no matter where we live.

It is this same advantage—the advantage, so to speak, of having a choice continent all to ourselves—that has spared from many of the ridiculous and useless quarrels, which have always been so prevalent in Europe.

When we have been forced “to adjust matters,” the traditional reputation of the Englishman, for being able to go right through with any situation, has triumphed.

Hoping to hear more good things from you soon,  
I remain, Your Friend, JOHN

No. 68.

### ESSAY B1

#### *Outline*

Main Theme.

Superiority of England.

Introduction.

“All men were born free, now man is everywhere in chains”—thus Rousseau; but not everywhere.

Main Part.

The English stolidness as opposed to hot-headedness.

Characteristics of English thinking.

Beauties of England—widely differing.

The countryside.

Conclusion.

The reason for England’s happiness to-day is that she is free, and every smallest component of it has his say.

[99, Oxford Rd.,  
Middlemore.]  
Nov. 5th, 1937

Dear André,

You say that most people consider their country the best in the world. It is only fair to say that they are perfectly justified in thinking so, but naturally I wholeheartedly disagree with them. I honestly think that, being the happiest country to live in, England is the best country in the world.

Rousseau, in his famous book “The Social Contract,” opened with “All men were born free; now man is everywhere in chains.” “Everywhere” is a sweeping generalisation, for obviously Rousseau had not made a stay in England. England’s greatest gift is freedom. Freedom personifies her actions, whatever she does. Even in the great War, she stepped in to save Belgium from going into bondage, which was what would probably have happened.

Freedom comes out in other ways. In countries ruled by a dictator with an iron rod, freedom has never been heard of. The citizens would be executed by thinking of freedom, because it is against the principles of a dictator. They may not speak against their rulers, either in public or at home. In England we may say and think exactly what we like, and may express our views in rhetoric, as long as we do not create a disturbance. It may be the same in Franco. I do not know. We may elect our government, and vote for any party under the sun. This is impossible in dictatorship countries.

Democracy breeds freedom, and England has the foremost democratic government of the world, with France as a runner-up. I am convinced that democracy is the only method of government which will solve the world's problems. Aggressive dictators do not try to do this. They just create more.

England is lucky in that she is an island. She has no frontiers to guard, like European countries. Admittedly it is by no will of her own that England is an island. If she had had frontiers to guard, her ownership might have been changed so radically as that of the boiling and seething hotch-potch of the world—Europe. It is our good fortune that we can, or could in the past, stay clear of the plots and counterplots of Europe. With the advent of the aeroplane, it is natural that the system of frontiers has broken down, and our country can now be as easily attacked as any other, within reason.

England can be looked at from another and more material aspect, the aspect of some very beautiful and widely varied countryside. With all the clashing marvels of the world, the English countryside has a soft and harmonious effect on a country-lover. I suppose every country feels the same about its countryside, but I doubt if it can compare with England, with her gaunt rocks and Scottish lakes, her mountains in Wales, and golden fields of corn, waving in the cool breeze of a Suffolk nightfall. Glorious indeed is the pageant of colour and countryside which is paraded before the Englishman's eyes, and he ought to feel thankful for it. I sometimes doubt if he is.

England's happiness depends upon a large number of things, but all these converge on freedom, and the ability to enjoy it. If the people under a dictator's rule were suddenly and unexpectedly freed, they would probably feel so much like a fish out of water that they longed to be back under their dictator, with his old familiar oppression, again. It is no use at all being free, if you have not the ability to enjoy it, and the knowledge of how to use it. But England has that ability and that knowledge, and uses them very sensibly.

Yours sincerely,  
ARTHUR.

138. For the undirected subject C the best candidates selected for Table II (5) were 41 and 95, whose marks are as follows:—

	<i>Examiner</i>	P	Q	R	S	T	U	V	W	Average
No. 41	Sense	11	12	11	9	10	11	9	11	10.5
	General									
No. 95	Impression	11	12	10	10	10	10	9	12	10.5
	Sense	11	11	9	8	11	13	11	13	10.9
	General									
	Impression	11	11	9	8	11	11	10	12	10.4

No. 41.

## ESSAY C

## A SCHOOL SPEECH DAY

For weeks beforehand, there has been excited chattering amongst the pupils. "What is it all about?" you may ask. You will soon receive an answer. "Don't you know? Speech day in a fortnight's time." Then you realise that the lusty voices you hear every day floating from the Hall, must be the various forms rehearsing the songs which are to delight their parents; that the speech day is the reason for the conversation often turning to white frocks and coveted books; and that it is because of this great day that the Head Girl is busy preparing her speech. The list of prize-winners is read out. The lucky girls reap the reward of hard working by being allowed to chose a book for a prize. The singing mistress rehearses the songs and brings them to a high standard; and the girls are given their positions for the great day. Then, at last that day arrives.

In the morning it is often bright. All the scholars attend the final rehearsal and put their hearts into bringing the songs to perfection. After rehearsing for most of the morning everyone returns home to spend the afternoon there. Then comes the evening. All preparations are begun and just as the white frocks are brought out down comes the rain, or else a fog. In thick dark coats or 'neath large umbrellas the scholars troop to the hall for this day for which they have been waiting for so long.

Inside the hall, there is more excited conversation. The parents, in duo course, begin to assemble. They try, though nearly always in vain to find their daughters in the sea of white below them. How alike the girls look! If Jean did not possess auburn hair Mrs. [Robinson] would never have recognised her daughter! Then finally the conversation dies down. The great occasion begins.

Having duly welcomed the guests, the chairman begins. During the course of the programme the scholars raise their voices and sing the songs which they have been practising for so long. Then the climax of the evening is reached. The "guest of honour" as we might call that person distributes the prizes. The receivers are given due applause and proudly receive their rewards. Then comes the speech. Sometimes it is about school and education, once it was about travel, and once about the past, present, and future. Whatever the subject, every one listens with great interest, determined to accept and use the advice and criticisms given.

After this speech, the votes of thanks are proposed and seconded and after the singing of the National Anthem, a very enjoyable evening breaks up.

The school speech day that will probably linger most in my memory was the one held in our own school hall a few years back. By means of microphone and loud speaker, the speeches were relayed from the audience in the Hall to another audience in the Gymnasium. This in itself was memorable because it had never been tried before. It was a novel experiment. I also remember it well because we, in the "gym," were allowed the first glimpse of Miss Y. who was later to speak to us.

This particular evening closed in a most original way. The lights fused and the building was plunged in darkness! Although it was a pity to conclude such an occasion in this way, although accidentally, the

failure of the lights caused great excitement, and was talked about for a long time afterwards.

The school speech day is always looked forward to and always remembered for long afterwards. The occasion serves to bring together the mistresses and the parents of their pupils. Careers are discussed and advice, where it is needed, is given willingly. On this day too, old girls, some of whom may have left the school long ago, return. Once again they meet the staff and talk with them and by attending Speech Day they see the pupils who are following in their footsteps and doing what they did years before.

No. 95.

### ESSAY C

#### A SCHOOL SPEECH DAY

Once every year, in the Autumn Term, the school is thrown into turmoil. The cause is Speech-day. The pupils' feelings about Speech-day vary with their age. The little ones are very excited and wander about with an air of importance. The older ones await the day with, perhaps, not quite so much excitement. They enjoy showing off their school to visitors but they have an eye to the time given up for Speech-day as well.

For weeks beforehand the pupils who take music are rehearsing songs and hymns to be sung on the day. People who take art are preparing designs and pictures to be shown on the day, and the needlewomen are making clothes and other useful things which will be shown in the Domestic Science Room. But during the last week everything is hurried up. Girls who are in the Chair are hauled from their lessons to have last-minute rehearsals. Drawings are hung in the Art Room. Last-minute glances are taken at needlework to make sure that there are no ends. Then at the end the whole school meets in the hall to make sure that everyone knows just what she has to do and where and when she has to do it.

But meanwhile another side of the preparations has been taking place. Invitations have been sent out to parents. People have been commissioned to serve tea in the Dining Room, and some forms have given up their gymnastics or games in order to carry chairs to the hall. These are taken from the loft generally but on the last morning nobody is surprised if she has to sit on the floor in her form-room because her chair has been taken. Speech-day is a very important event among the parents and therefore every inch of space in the hall, and every chair, is used. On this great occasion pupils sit on chairs also, instead of sitting on the floor.

White is worn by all pupils on this day. Every girl rakes out her white dress, which has been stuffed at the back of the wardrobe since last Speech-day, and tries it on to see if it can possibly be lengthened or whether she ought to have a new one. When she has decided brown shoes are given an extra polish, light stockings are inspected in order to make sure that there are no holes, hair is given a really good brushing and then she feels that she is ready for the great day.

At last Speech-day arrives. The school musters in the hall. Mistresses are all wearing their gowns, which make a sharp contrast against the white of the school's clothes. The parents begin to arrive next. At first they arrive in two's and three's but then the number increases until there is a stream of people pouring in. Pupils whose

parents are coining watch carefully for them while the parents scan the lines of pupils for their daughters' faces. Then the more important visitors go to their seats on the platform. These are the Head-mistress, the Chairman, the person who is to present the prizes and members of the Education Committee. Then the Chairman announces the first song, the choir clear their throats and Speech-day has begun.

The afternoon passes. There are speeches and songs, songs and speeches. Then the person who is to give the prizes rises and prize-winners get into order. As they go up to receive their prizes the rest of the school claps and tries to look as important as the prize-winners. They merely succeed in looking as if they could have won prizes if they had tried. When the prizes have been received the house-captains and form-captains go up to receive the cups and trophies which their house or form has won. Cheering never ceases here. Then the lady or gentleman who has given the prizes makes a speech. Those are generally very interesting. But perhaps the biggest cheer is for the Chairman when he announces that we may have an extra holiday.

Now at last Speech-day is over. The next day it is talked about for a little and then everything falls into its usual routine until the next year.

139. For the directed subject Cl, the marks of Nos. 25 and 68 were as follows :—

	Examiner	P	Q	R	S	T	U	V	W	Average
No. 25	Sense	13	11	11	10	10	12	8	11	10.75
	General Impression	11	10	8	9	10	10	10	8	9.5
No. 68	Sense	7	14	12	12	12	12	12	12	11.6
	General Impression	8	14	11	12	12	12	12	12	11.6

Although the average mark in each case is distinctly above the median of our scale (8), No. 25 has received three 8's from different examiners, and No. 68 has received a 7 and an 8 from the same examiner. In the case of No. 68 the opinion of Examiner P differs strikingly from that of his colleagues, whose concordance is exceptional.

No. 25.

#### ESSAY Cl

[11, Derby Road,  
Northmore.]  
July 8th.

Dear Jean,

I was very sorry that you were ill, and unable to attend the speech-day yesterday. I am just writing to tell you all about it, as I expect you would like to know.

The Hall was crowded, as usual, with parents and friends, some who had probably come for the first time, others who had been for four perhaps five years. But, from the various comments I heard, I think that everyone enjoyed the afternoon.

The afternoon before as you know we held a full rehearsal, and every girl knew her place beforehand. At a given signal there was silence in the stone hall, and we marched into our places, the organ playing, and the mistresses in their gowns. A short time after we had been seated, the platform party came in, the Canon as usual taking the chair, and the headmistress telling the others where to sit. The lady who was giving the prizes sat to the right of the Canon, and the headmistress to her right, just the same as it has been always. At this moment we were all standing—parents and friends sat with the platform party, but we remained standing and sang the school song. I think we sang it quite well for Miss X. looked pleased.

The Canon then rose and gave the chairman's address, this was not too long, and he then announced that "the school Choir would sing a three part song." The choir rose at the signal, and Jean, after all the practicing of getting up all together we did it. That was a good start! The song went off well, even the poor seconds were heard occasionally too. When we sat, the Canon announced that "the head mistress will now give her report." Miss Y. rose and thanked Lady A. for kindly coming and giving the prizes, and then went on to give the usual account of the school work, and new building, and efforts to gain enough money to have a swimming pool. This caused a good deal of laughter, and applause. The games report was good, also the examination results.

When Miss Y. had seated herself, amid great applause from her pupils and the parents, the Canon rose again, and announced that the school would sing in unison. This song went off well, and I think everyone enjoyed it as they gave us a hearty clap.

Lady A. then gave us her speech and really it was topping, I have never been so interested in any speech given at an afternoon of this sort. All the parents thoroughly enjoyed it. You will probably see the whole of her speech in the local paper this week, so I will not spoil your enjoyment by telling you about it.

Lady A. sat down amid our famous Kentish Five, and the Canon had to hold up his hands to silence us. The choir then sang a song and it was well liked, I don't think we sang quite so well as we did the morning before, but Miss Y. seemed to think we did quite well.

Next came the great event, the distribution of prizes and certificates. Lady A. did this charmingly, while the admiring parents and envious girls looked on, perhaps the girls who looked most pleased, were those who had gained school leaving cert: !

After all the girls had regained their seats and the applause had finished the choir sang their third and last song. The votes of thanks followed this, and everyone praised the gracious way Lady A. had distributed the awards and had said a pleasant word to everyone.

After this there was a short interval. The platform party came and sat in the front row of the audience and the curtains of the stage were pulled across, and everyone waited for the play to begin.

As you know, this year there were two plays—a Latin one and a sketch of two girls from "School for Scandal." First came the Latin play, of course no-one understood it so the acting had to convey all that was going on to the audience, which it did. I and everyone else thought they acted extremely well. The costumes were very clever too.

The other play was very different from the first. It was a scene in the house of Sir Peter and Lady Teasle. They were having a quarrel

and this was acted very well. In parts they made every one roar with laughter. The two dresses were exquisite. Lady Teasle, had a beautiful pink tapestry frock on, shot with gold and blue, with lace round the neck and cuffs, she carried a pretty fan. Her hair matched the costume too. Sir Peter's coat attracted much attention, it was in purple velvet with a very charming lace cravat, and she wore breeches and black patent shoes with large buckles. She also wore a wig and carried a black walking stick with a silver knob. Everybody liked this sketch, and greatly praised the producer. This sketch had a great applause and after the curtain went down there was a buzz of conversation. The platform party then left the hall, but the parents were asked to remain seated until the girls had marched out.

In the stone hall once again, there was the usual rush for hats and coats and another rush for the rather small exit.

I must close now after that rather rambling account. I am hoping to see you back at school soon, as I am sure you do not want to miss examinations !

Your affectionate friend,

[CAROLINE.]

No. 68.

ESSAY C1

[99, London Road,  
Westmore,]

Nov. 12th 1937.

Dear John,

Being absent, you have missed an extremely interesting school function, the most important function of the year, as our respected Headmaster has once again reminded us. I am convinced that he reads the same notes on speech day every year, although he expects us to laugh just as heartily and listen just as attentively.

At about a quarter to two, the usual mêlée of bored boys and over-excited parents arrived at the gates. Eventually the porter, resplendent in a new uniform with red collar and cuffs, opened the doors with a dramatic flourish. In trooped hordes of parents, as usual mostly mothers, all intent on getting front seats, showing a decided preference for those reserved for Governors and Friends, from which laudable intention they were politely but firmly told to refrain.

With much clattering of bathchairs, and learned conversation, the Master and Governors of the Worshipful Baby-Linen Company entered, the more honoured sitting on the platform, and the rest in the seats allotted to them. The Master then arose, and proposed a vote of thanks which was completely inaudible, following the usual custom of votes of thanks. Amidst rather subdued clapping from boys and hysterical cheering from parents, the Headmaster rose. He made a long speech about the school's activities, lasting a quarter of an hour, politely stressing every success, and omitting every failure. He sat down, and almost simultaneously, the Worshipful Master he rose, amidst more clapping. He made a long speech about the deplorable lack of observation in the world today. The parents pretended to understand it, but the boys did not even reach that height. I forgot to say that before this there occurred that deplorable institution, the prize-giving.

Then, with clapping from the parents only, and a strong under-current of booing, sad to relate, Mr. [Weston] appeared with every

indication of haste, including a wrenched collar. He proceeded to wave his arms about in the approved conducting manner, making his face even redder than usual with the unwonted energy. The choir sang about half a dozen songs including of course that famous song about "Sceptred isles," and "seats of Mars," which turns up every year or so. Another horrible catcall was a very unmusical round, appropriately entitled "Busy, curious thirsty fly," which, as far as music and harmony are concerned, was nothing more than a farce.

A French speech, with no explanation, followed, ably spoken by [Jones], although nobody understood it, except the parents, who hid complete ignorance under a mask of interest and mild criticism, as if detecting some fault in [Jones's] accent.

There was a heartily applauded declamation on "Superstition" listed on the programme, but as the declaimer had a very largo mouth and a very small voice, I had to content myself with looking at his mouth or else read my prize, which I was loath to do. My mother afterwards informed me that it was an excellent declamation, so I suppose it must have been.

The choir then again appeared on the platform, each member taking up his original position. How they all managed it, I cannot imagine. They sang "Carmen —" with great gusto, as they had the rest of the school to back them up, and lastly came the National Anthem, which Z.<sup>1</sup> refused to sing, on the grounds of being unpatriotic to [his country]. The boys immediately disappeared, the parents remaining behind to gossip, drink stale tea and eat raw buns. The end of a perfect day!

Yours sincerely,

ARTHUR.

#### (B) EXAMPLES OF ESSAYS REGARDED AS JUST "COMPETENT"

140. It will be remembered that the Sub-committee decided that the median mark C of our fifteen-point scale, A<sup>+</sup> to E<sup>-</sup>; should be used to designate "competency" (see para. 29 above). In an examination designed to test the possession of a "utilisable skill" in English this would be the pass mark.

Among the "best agreements" to which our statisticians drew attention are the marks for General Impression for two essays on the subject C1, by Nos. 34 and 70. We add here, as in other instances, the marks for Sense.

Examiner	Essay-subject C1								Average
	P	Q	R	S	T	U	V	W	
No. 34 {	Sense	5	8	8	8	8	8	8	7.6
	General Impression	8	9	8	8	9	8	8	8.25
No. 70 {	Sense	7	8	8	8	7	7	8	7.5
	General Impression	8	9	9	8	7	8	8	8.1

It is by pure chance that the one essay is by a girl and the other by a boy.

<sup>1</sup> The name shows that Z. was a foreigner.

No. 34.

## ESSAY C1

## A SCHOOL SPEECH DAY

As you know the Speeeth Day was yesterday. Probably you were pleased you missed it, thinking it would be a bore having to listen to a lot of speeches. You were wrong then, if you did, because wo had a jolly good time. The person who presented the prizes (Miss X.) was very nico and said quite a few funny things which made the school laugh heartily. When wo arrived all the mistresses were there in beautiful frocks and wearing their black gowns. Not many of the parents had arrived, and we amused ourselves trying to guess which girl belonged to which parent as they came in. It was rather fun, although our guessos were probably quite wrong. All the lights were on, and the prizes wore piled high on the table on the platform. The scenery was very pretty—flowers and trees and bushes, like the usual colourful show found in a cottage garden in summer,—although it is now in the middlo of Octobor ! It all looked very sweet, though, and when the staff came on the platform to sit down, they all had beaming countenances.

There were the usual introductory speeches, intermingled with one or two songs sung by the Upper and Middle Schools. One was a lullaby "Sweet and Low," and was received with great enthusiasm by the parents and staff.

The prizes were then given out, and Miss X. shook hands with each of the girls in turn, usually having a word of greeting or encouragement to say to each of them. [Diana], of course, tripped up on the last step and came floundering and stumbling on to the platform, but managed to look calm, though ruffled, as she stepped forward to receive her books !

When she had finished giving prizes, Miss X. delivered a little speech. Thank goodness, it was not so very long, and so we were able to appreciate it. She said that we must go on working hard, and that we should have to go on keeping the good standard of work work that we have, judging by the results of the exams, and the prizes. She realised, however, that we did not want a lecture, and so she gave a few illustrations of varions funny incidents that had occurred in her life, talked of one or two other prizegivings she had attended, gave us a little advice on work and play in general, and then sat down, amidst furious applause. One of the reasons that she was liked so much was, I think, that she was reasonably young and pretty, and so looked very different to some old proffessors that we have had.

We had a few parting words from one or two of the governors and councillors, and then presented the principle people with bouquets. These certainly were beautiful. After singing the School Hymn, we left, all feeling very pleased with our evening. We did not get home too late, and I feel sure you would have enjoyed it, had you been there.

No. 70.

## ESSAY C1

[33 Cambridge Road,  
Southmore.]

9th June 1937.

Dear Frank,

I hope you have recovered from your unfortunate illness which prevented you from attending school on Speech Day. However, I now endeavour to reproduce in writing what happened on that day.

On the previous day, we were informed by the Headmaster that we should not assemble in the Hall at the usual time, but that we should wait until we were summoned by the prefects. From nine o'clock until the time came for us to go down to the Hall, we amused ourselves in numerous ways. Some played "noughts and crosses," others played other games, and a few worked. Then came the prefect. So with one accord we rushed to the door; this was undoubtedly because we were so eager to listen to the speeches. Or was it because we wanted to become comfortably settled in the Hall. In a very disorderly order we went down the staircase and into the Hall which was all ready for our reception. We sat down and immediately had to rise again because the Chairman of the Board of Governors, his wife, and other Governors of the school came in with the Headmaster. After this party had established itself on the platform we again sat down, this time for the whole morning.

The Headmaster began the proceedings with a summary of the year's work which definitely gained the approval of the Governors. The Chairman's speech was the next item on the programme. This speech was similar to last year's speech, which you heard, the essential detail, the opportunities for [Y] boys in "the outposts of Empire," being the same. The Chairman introduced the distributor of the prizes, thus ending his speech. After the Prize Distribution, the distributor made an enjoyable speech which was appreciated by the whole school. A Vote of Thanks was proposed by one of the Governors and seconded by the Secretary of the old [Y-ians] Society. Cheers were called, and given, for the distributor. A half holiday was proclaimed, and the meeting broke up with school over for the day.

I hope you will soon be well enough to return to school although I expect you feel differently about this.

I remain,

Yours truly,

[F. H. SMITH.]

141. Both No. 34 and No. 70 make some mistakes in spelling. (More distinguished people do the same.) But it is clear from the marks of the examiners for General Impression that they regard these essays as showing a "competent" command of English—without any exceptional qualities.

We do not attempt here to analyse the essays further as this would be to import personal judgments other than those of examiners into this section of the Report.

#### (C) EXAMPLES OF ESSAYS IN REGARD TO WHICH THE EXAMINERS DIFFER WIDELY IN OPINION

142. There is some interest in considering a few specimens of essays in which the agreement of the examiners was worst. In those selected the differences between the examiners in their marks for Sense and General Impression vary from 4 to 8 points on our 15-point scale, which would correspond numerically to

the large differences of from 26.7 to 53.3 points on the 100-point scale in general use.

143. The following are the marks for Sense and General Impression for four essays on subject A :—

Examiners	P	Q	R	S	T	U	V	W	Range
No. 9 { Sense	7	2	11	5	6	11	10	5	9
	General Impression	7	7	10	5	6	11	10	8
No. 22 { Sense	12	5	10	8	9	13	9	11	8
	General Impression	12	9	10	9	8	12	9	10
No. 45 { Sense	11	9	12	8	7	7	7	10	5
	General Impression	11	9	12	8	6	8	7	11
No. 80 { Sense	7	8	11	5	9	12	8	8	7
	General Impression	7	8	11	5	8	12	8	7

144. We quote below the essays of Nos. 9, 45 and 80.

No. 9.

#### ESSAY A

At any school, you will nearly always find that the girl or boy who shines above the rest at games, is revered more than their fellow-students who can do their lessons better than anyone else in the class. At school matter triumphs over brain. Your status is decided by your skill at games not by your skill at lessons and although the boy who wins a pile of books at speech day gets a good clap and is called " quite a good chap," it is the boys who gets his colours that wins the most applause from boys and parents.

After hard work in the class-room, girls and boys are only too glad to get out on to the grass and to forget a badly written exercise that has just been returned. But a fumbled catch at cricket remains in the memory much longer than the bad exercise and scathing remarks of the master or mistress.

In the corridor is a large notice-board, the green baize covered with papers. A crowd of girls are round it, very eager to see who has got into the first eleven and when the next practices are. A french lesson should begin when the bell rings but there is too much noise at the notice board to hear it. A lot of girls there are weak at French and they must pass General Schools to get the job they want, but it becomes of secondary importance when the games' board is full.

Of course, games are much more pleasant than ordinary lessons. Even the " rabbit " at games admits that she enjoys being out first ball and watching the rest of the game much more than concentrating in a form-room.

English youth is naturally very fond of the open air and full advantage of the games period is taken. Take this keeness and also the keeness of the young games master or mistress and you can see that it is keeness and the desire to do well that urges on the would-be champion at games.

At the kindergarten school in games lessons, the first thing we learn is that you cannot play games without dividing at the least into two sides called teams and thus from the beginning, team-spirit is brought up against us.

Even if you are very small and only playing with a big rubber ball, to know that you must catch it every time it is thrown to you and to have missed a catch means that you have let your side down. Then you feel that you do not want to play any more, but by the time you have lost your first school the only desire is to get into the school team because you know that you will never let your side down.

Work in the class-room is so individual. The only person who is let down if you do not work with a will, is yourself. There is no one else to be disappointed except your parents who are sorry that you have been disappointed yourself and the mistress at school, although I find she is rarely given a thought.

In both work and games we have a goal and that goal is to win what we want. In work if we fail, we say we will do it next time but in games if one fails the rest fails.

There are too many games lessons and matches in school life. Yet in actual fact, if I had to miss any part of them I should feel both indignant and disappointed as everyone else would. But to miss lessons because of a lecture or concert is bliss.

When it comes to everyday life, and as the bell has just rung and I can hear the crowd round the notice board outside the door, I can only wish to be out there, and think there can never be too many games in school life.

#### No. 45.

##### ESSAY A

Main Theme	All work and no play . . .
Introduction	Games and the pupils future.
Main Parts	(a) Physical effect of games on pupils. (b) Mental effect of games on pupils. (c) Association between master and pupil. (d) School life without games.
Conclusion	Games should be upon an equal status with learning in the school curriculum.

There are those who say that children go to school to acquire knowledge and not to play games; such people do not trouble to look before they leap, for if they did, they would not despise games as not coming under the heading of knowledge. Nowadays, if one wants to be popular among the ordinary people, one must have an ordinary knowledge of the present sports situation as well as the present political situation.

If in an ordinary school the pupils had no chance of physical exercise at all, their mental outlook would suffer for to be physically fit is to be mentally fit and vice-versa. Thus, it can be seen that some form of physical exercise is necessary to the pupil. In addition to the other advantages, it has the effect of bettering the pupils health.

If the effects of cooping a child up in a room full of books for a considerable period, could be observed, I feel sure that whatever form they took, they would be detrimental to the health of the child because the child would either read and read, or else be miserable all the time and in either case the effects upon his mental health would be bad.

Games are not only absolutely necessary for the good of the child's health but also form, usually an effective link between master and pupil: this is because it is usual for the master to arrange the games and control them, while the pupils play them and thus the master has to have a

knowledge of the game to control it, similarly the pupil, to play it, and thus a common note is struck between them.

Without games it would be putting things mildly to say that school life would be intolerable, for it would be mental torture for the masters and both mental and physical torture on the pupils. The mental torture for the master would come from the fact that the pupils would be thoroughly miserable and unable to absorb knowledge with their customary readiness and this lack of interest would make itself shown in the ill-temper of the master. The pupils would then suffer for the temper both mentally, by the impending physical punishment, and physically, by the punishment. From this it can be seen that the indirect cause of all this trouble is a lack of physical exercise for the pupil, in other words, little or no games.

From the preceding paragraphs it can be deduced that for the good of master and pupil, in fact, all those connected with the scholastic establishment, the place of games or physical exercise in one form or another must play just as important, or nearly as important a part as the acquiring of academic knowledge for it is obvious that with games and learning it is a case of—"Together we stand, divided we fall!"

#### No. 80.

#### ESSAY A

How the word "Sport" makes the heart of many an English Schoolboy beat faster. In the Summer, cricket and tennis; in the Winter football and rugby, all of them good healthy games which have helped a great deal in making England such a fine country. What enjoyment the youthful schoolboy finds in imagining himself as a Herbert Sutcliffe, Patsy Hendren, or Dixie Dean; and one may say that the greatest follower of the "Big" Sport is the scholar.

Undoubtedly games play a great part in School Life, and it is expected that they should do. The boy who sits in a desk crouching over books for a great part of the day is in need of a regular relaxation from the old routine. Some form of this is provided by gymnasiums, with which most of the schools are now equipped. Even the smaller and ill-fitted schools carry out their regular "drill."

Games should not, however, interfere with the serious work, but it is sometimes unavoidable for ones mind to wander round the latest score of the Test Match, and the "come back to earth" when the master asks an unfortunately timed question. But there is no doubt that in many cases the sport enables one to think quicker, and also gives a very welcome break to the master.

In recent years the importance of school sport has been impressed upon the minds of people by means of the newspapers. It is considered that the only way to have a strong and healthy country is by constant physical training. Unfortunately this policy has not been followed by Great Britain until just recently, and consequently we are rather behind several of the European Countries. Much of the drill of the European countries, in some cases compulsory, is carried out with a "clockwork precision" which is woefully absent in this country.

However, the lack of the games is now well on the way to mend, and as well as the schools' physical drill, the unemployed are given similar chances to keep fit even though lack of work may cause drooping spirits.

But even through these hard and trying times the old British sport and sportsmanship will continue unforgotten, and the names of great cricketers and footballers will hold a very high position in the minds of schoolboys.

145. The following are the marks for Sense and General Impression for four essays on subject A1 :—

Examiners	P	Q	R	S	T	U	V	W	Range
No. 46 { Sense	1	2	8	2	2	5	9	5	8
	3	6	6	2	4	7	7	6	5
No. 74 { Sense	9	5	7	4	9	8	5	11	7
	8	8	7	5	9	9	7	12	7
No. 85 { Sense	11	13	12	7	4	9	11	12	9
	11	13	12	7	5	10	11	11	8
No. 91 { Sense	6	7	11	10	9	13	9	14	8
	6	6	8	10	9	12	8	13	7

146. We quote below the essays of Nos. 74 and 91.

No. 74.

### ESSAY A1

#### SHOULD GAMES BE REGARDED AS A DUTY OR A PLEASURE?

This subject would be a very popular one at a school debate as most girls and boys have very strong opinions on the subject. I think more people regard games as a pleasure than as a duty because they enjoy outdoor games. Some people find pleasure in some duties and they may think of games as duty and still enjoy playing them.

Games are always included in the school curriculum nowadays. Most girls seem to regard this period for games as a pleasure, when they do not have anything difficult to do, and where not so much concentration is needed as in the form-room. Those who are good at games most certainly do look upon games as a pleasure but it is doubtful whether those who are not good do as well. I am not very good at games but I do enjoy them and find them a pleasure as much as those who are good at games. It seems to be such a relief to leave the form-room and go out to play games more so in the summer than in the winter. I think most people enjoy the summer games more than the winter ones. Rounders, tennis, swimming and cricket all seem more enjoyable than the winter games. This is, of course, a matter of opinion. Although I enjoy summer games, I think there is nothing more exciting to go out and play hockey when it is very cold and frosty ; as the cold wind makes you feel so fresh after playing for about half an hour. I personally do regard games as a pleasure because I enjoy playing them.

Many people may regard games as duty because they think it is good for them and makes them more healthy. Some people like one game more than another so that those games that they have to play which they dislike they would consider as a duty. Different types of people may consider games as a duty. Those who are fond of drawing and sketching or fond of reading and studying may consider that it is their duty to play games and may not find much pleasure in playing them. Those who always like to be out of doors or playing games would naturally look upon games as a pleasure. This shows that you

have to think of the types of people before you can decide whether games should be regarded as duty or a pleasure. People who are fond of drawing and studying may also find games a pleasure but I think it is more likely that they would find them a duty.

People of the outdoor type who play games for the pleasure of playing would naturally play better. I always think that you naturally do a thing better if you are enjoying it. Therefore, the people who play games because they find it a pleasure will play better than those who consider it a duty.

You have to consider the question of the time when you play games. At school you have to play games and those people who find it a duty would not like this. Out of school if you want to play games you may, but if you do not wish to you are not forced to play. Those people who play games out of school naturally find it a pleasure or else they would not play.

As so many people play games out of school hours I think that more people find games a pleasure than a duty. This of course is to an advantage as naturally games will be played better; therefore games should be regarded as a pleasure rather than a duty.

#### No. 91.

##### ESSAY A1

###### SHOULD GAMES BE REGARDED AS A DUTY OR A PLEASURE ?

If games are to be regarded as a duty or a pleasure there must be some invincible reason for the choice, and in choosing certain important factors must be taken into consideration.

If they should be thought of as a pleasure certain questions must be solved. The aim of games must first be taken into consideration. There are various reasons why we play games and the main one is in order to obtain relaxation. The game may be a strenuous one but it has necessitated the use of channels of the mind not employed in every day incidents. From this relaxation the development of another side of our nature is in progress. We learn the sporting spirit of give and take, and we become fairer in our judgements, therefore we are more suited to be citizens of the community. These games usually take place in open air and as they all involve swiftness and precision those playing them develop a higher standard of physical fitness.

Despite the encouraging results of taking part in games, they have a definite place in community life and are not to be regarded as of primary importance. They are essential but cannot be thought of as more important than the every day pursuits of working to obtain a living and they should be confined to the times when work is disregarded.

The greatest joy in games comes when the particular game is first mastered. As efficiency is gradually acquired, the game adopts a more serious aspect. Then thoughts of proving one's ability are entertained and competition takes an important place, but despite this there is only a small section of the community who plays games professionally. The others regard them as being something to enjoy and as being superior to any indoor amusements, especially during the summer months.

When regarded as a duty yet another aspect of them is entertained, that is of the results of playing games. Then we find that people who play become strong and athletic, and have more resistance to illness and can therefore work better. The control of growth, of straight

limbs and square shoulders can be assigned to games. Therefore instead of a nation of physically unfit people, we get an A1 nation of strong supple men. During times when political disturbances are felt, the health of the nation must be of a high status in order not only to defend their country if need be, but to command the respect of others and to be able to make political agreements with the knowledge that their will be no hesitation and that one's country will follow one.

Another and important duty is that to posterity. We owe it to the nations to come to be of a high standard of good health ourselves so that they may be the same. Games should therefore play an important part in attaining this standard.

Whether thought of as a duty or as a pleasure we can see the essentiality of games. They should be regarded, primarily as a duty but as they give pleasure, as a pleasant duty. They should be enforced as long as they keep their correct place and the youth of the nation should, by being taught how to enjoy them, learn a true standard of their value.

147. The following are the marks for Sense and General Impression for three essays on subject B :—

Examiners	P	Q	R	S	T	U	V	W	Range
No. 12 { Sense	10	7	8	9	10	9	5	14	9
	General Impression	10	8	9	9	9	5	13	8
No. 77 { Sense	11	11	11	9	12	12	6	13	7
	General Impression	11	9	7	10	12	12	6	12
No. 84 { Sense	9	8	11	8	12	15	8	14	7
	General Impression	11	10	11	9	13	15	9	13

148. We quote below the essays of Nos. 12 and 84.

No. 12.

#### ESSAY B

This sceptred isle, this earth of majesty.  
This fortress built by nature for herself  
Against infection . . .      *Shakespeare.*

There are still many Britishers who strive to regard England to-day as the nation which leads the world and which all other peoples look to for guidance and with respectful envy ; and they are truly noble people who would, in their imagination at least, see England as Shakespeare saw it. And although we who inhabit this tiny island are but a thirtieth of the population of the world, and this isle itself constitutes but a four-hundredth part of the area of the world, we are, I say, still the greatest nation. Those who dispute this, do not know us—yes us, for I am a Englishman, and as proud of it as of anything else. England still leads the world in the things which are right, worth while, and just in life. The things which do not matter she does not, however, ignore, but tries to improve them, remedy them substitute good things for them. It is a great task. England has always succeeded, is succeeding, and will succeed. Her success in such a task is inevitable.

Many childish countries have with such petty things as embargos, threats, warnings and accusations tried to prevent us—they even dared to try to crush us! So mighty were their arms. But they will never cause the virtues of an Englishman to waver.

Throughout the centuries, since long before dictatorships or swastikas were thought of, England has been steadily, continuously establishing her present position. She has never tried to conquer the world in a fortnight, but always striven to reform the world—for ever.

Volumes and volumes and volumes of literature has been written in praise of the so-called good Old England, of the gallant sailors of the sixteenth, and of our soldiers through the ages. These could only have been of the one Great Country.

Even with the advent of comparatively modern monsters—those huge grey ships, and massive torpedoes, and submarines, and aeroplanes, and airships, and death-rays—yes, even with all these and others, England is not moved.

She abhors these things, but still is as calm as ever.

She adapts herself to the new conditions—for any sensible state must do this—and at the same time does the job well, but at the same time never ceases working for the good, not only of herself, but of the world.

This is the thing to be remembered. The minor states of Europe—they may have armies of millions and air forces unimaginable but they are still only very minor states—these states work only for their own benefit.

They are greedy, selfish; want everybody else's territories, money, and goods, only to squander and despoil when they have them. But England is the opposite and as opposite as anything could be.

A small but infinitely great country as beautiful as to be unique in the variety of her scenery, with her every member conscious of the glory of his fatherland, and of its unparalleled record of good work towards the perfecting of all the lesser nations (who do not even know themselves what they are really trying to do,—or what they *will* do when they have done this)—this is the England which exists TODAY, if only all could realise this. We do not more than necessarily boast of our past and future. We broadcast our glorious Present.

#### No. 84.

#### ESSAY B

"This emerald isle set in a silver sea," said Shakespeare, speaking of his native land, England, and echoing and re-echoing down the ages have come stories of the pomp of her kings, the glories of her triumphs and the might of her people. Her throne alone has stood secure, since the days of King Alfred, and still through the centuries come her kings with stately tread and sweeping velvet. The steady beat of the drums of her unvanquished armies resound along the endless path of time, and lead their people to glorious victory.

Yet what is all this might, this grandeur, without something deeper? Has England only outward splendour? No, deep in the works of her poets, her authors and painters, treasured in the heart of her people is something without which she would be nothing: the faith which keeps her true to her kings, the love for her fellowmen, and the humility which still allows her to acknowledge that there is one King, whose beauty

love and glory far surpasses the greatest of her kings. Without this she would be nothing ; she has so much pride, and yet the worship of a Great God can make her feel so very humble. In the glory of her cathedrals, in the solemnity of her churches in the heart of every true Englishman is found the beauty of our religion.

The rugged hills of the north lift up their heads to the heavens, and the quiet green meadows of the south bathe in the caressing sunlight. What country can give anything so fair as our countryside in the spring ? Around one is the singing of the birds, the rushing of tempestuous streams which hurl themselves from the hillside. Above one is a dappled sky, where a sickle breeze chases the fleeting clouds. This is the country where men are born, poets, statesmen, soldiers and sailors. The sailors of the west sailed the seven seas, and brought home to England fame and riches. With the help of a few English ships, and stout hearts to man them, they brought the might of Spain's Armada to nothing. Where were the Spanish admirals when Drake's drum thundered across the seas ?

The mighty Napoleon, driving all before him was no match for England. Nelson in his Victory, saved the day and soon Napoleon, the proud, the mighty, began his humiliating journey to Elba.

Again and again, with glowing pride the Englishman hears stories of his country's glorious history, and as the armies tramp along, and as the kings come swooping by, yet still he hears the grander, nobler voice of the one King.

#### 149. The following are the marks for Sense and General Impression for four essays on subject B1 :—

Examiners	P	Q	R	S	T	U	V	W	Range
No. 40 { Sense General Impression	8	8	8	6	5	10	5	3	7
	8	8	7	6	4	10	6	5	6
No. 46 { Sense General Impression	6	9	12	11	10	9	11	8	6
	3	10	6	8	5	10	10	7	7
No. 58 { Sense General Impression	4	9	7	5	5	10	12	6	8
	7	9	7	6	6	10	11	9	5
No. 88 { Sense General Impression	7	11	8	10	12	10	7	11	5
	6	11	7	10	11	9	6	11	5

The essays of Nos. 40, 46 and 58 are all regarded by four or more of the examiners as below the border-line of " competency " (8), in either Sense or General Impression or both ; but No. 40 has received a mark of 10 from one examiner in both categories ; No. 46 has received four marks of 10 or more for Sense and three of 10 for General Impression ; and No. 58 has received two marks of 10 or more in both categories. The opinions of the examiners in regard to these essays are clearly irreconcilable.

150. We quote below the essays of the three candidates above referred to :—

No. 40.

ESSAY B1

[Rose Cottage,  
Chester Road,  
Southmore.]

Dear Marie,

In your last letter you praised France so much, that I can do no other than show that you are wrong in saying France is the best country in the world.

Although England is my own country, I'm not going to praise it because of that, but because there are so many things that make it stand out above other countries and lead them. Surely if England were not the leading country it could not defend and please such a vast widespread Empire as it does.

The English parliament is a model to other countries for it thoroughly goes into every detail of any question or misunderstanding which may crop up between parties, two firms or even the employees of a firm over the question of wages and hours etc.

English politicians are always watching other countries and when a difficulty arises there they are only too willing to help straighten it out, although they are very careful not to be drawn into any war.

There are very few changes in the parliament, changes only taking place when a leader or any member dies or retires or at the general election when it is compulsory for everyone to be re-voted.

Everyone can go to bed at night with a safe feeling and not worry whether anarchists are going to rise up in rebellion during the night, for everyone loves and respects the Royal family and monarchy.

It is also quite safe to go out for gangsters are practically unheard of in England except through newspapers and films of foreign countries. Of course I'm not saying England is absolutely free from criminals because it is not, I do not think that any country could be but Scotland Yard is very particular about the crimes committed in the British Isles and a strong record is kept of every criminal. Some of the most up to date methods are used and very few people escape from "the clutches of the law."

The country itself is very pleasant to live in for there are no dangerous wild animals and reptiles to watch for. The atmosphere is pleasant and people willing "to help lame dogs over styles." In summer it is never unbearably hot and in winter it is never so cold that people get snowed into their house and frozen to death. Although England has her share of rain it does not rain every day and all day while on the other hand it does not suffer from drought.

Many pleasant walks can be taken in the country where numerous wild flowers grow in their natural surroundings along the hedgerow, in woods or by the sides of streams. Whichever month of the year is chosen for a country walk there is always something to see, and there are always birds who chirp and wake one up on a summer morning or visit daily for crumbs during the winter.

Being an island there is plenty of coast and fortunately there is a tide all round the coast which means that the beach is washed twice daily and therefore are quite healthy for the dirt gets carried away.

Poets all through the ages have been fond of praising England for

her beauty and power. It is not only parts of the country that are described in verse but statues, kings and towns.

One poet wrote of England

“ This Royal throne of Kings  
This sceptered island  
This earth of majesty this seat of mars  
This fortress built by nature for her purpose  
Against infection and the hand of wars.”

Another called it the “ fairest Isle ” and said

“ Fairest Isle all isles excelling  
Cradled mid the western seas.”

Many emigrants on far shores have pined for a sight of the coast of England and written :—

“ Oh ! to be in England now that spring is here ” realizing her beauty and what they are missing.

To many colonials returning to England for a holiday after longs years abroad in the backwoods and jungles of Africa, the bush of Australia, a lonely ranch in Canada or military service in India have welcome the sight of the “ white cliffs of Dover.”

This seems more of a lecture than a letter so I'll close before it gets too bad.

Please give my kind regards to your parents and write soon.

Best wishes,

from your pen friend,  
[Lucy]

#### No. 46.

#### ESSAY B1

England is without doubt the greatest country in the world. We have honour, prestige and tradition.

England has produced the greatest man in literature throughout the whole world—, in the person of William Shakespeare. France, Germany, Italy, have never produced such as genius as him—that is something which I think Englishmen should be proud to know. We have not got only one great man in literature either for names such as Milton, Shelley, Wordsworth might well we ranked with names such as Dantes, and it is an admitted fact that France has never produced any great poets.

To say, that England is better than any other country in the matter of scenery, is, I think an injustice, for while England does not better them in this respect I think it is certainly at par with them. In most countries such as Norway, Sweden, Switzerland, Italy, there is so much the same, over large areas, but in England we have variety in our scenery and “ Variety is the spice of life.” We are not confined to mountains, rivers, plains, for our beauty, but we have the whole lot moulded into one. In England if you want hills and mountainous scenery you go West and North if you want flat expanses of country—go east to the Fenlands, if you want gently undulating country—go South to, Sussex, Kent and Hampshire.

England is by far the safest country in Europe, this statement finds prove in the fact that many exiles from foreign countries on the

Continent find a refuge and haven in England and names like Haile Selassie, from Abyssinia, King Alfonson from Spain, instantly flash to ones mind, and Colonel Lindenburgh could not find refuge in the  $2\frac{1}{2}$  million square miles of the United States, but found it necessary to come to England and find peace in the Weald of Kent—at Sevenoaks. Therefore this proves does it not ? that England is safe and provides a hospitable hand to foreigners. Perhaps an even more striking instance of the hospitality and generosity of this country is the shelter afforded to the 4,000 Basque children who have just come from Spain to find shelter and peace from the bombing in their own country.

Perhaps in the matter of roads England is inferior to some countries such as America, but one must remember that England is an old country, the houses and roads have been built to suit the conditions of horse traffic whereas those of America have been built to suit motor traffic from the first. Many foreigners, especially Americans accuses England of slums and lack of town planning, but they forget that England has taken about 200 years to make her towns and cities especially the manufacturing towns such as Manchester, Leeds, and Birmingham, but most towns and cities on the Continent of America have been created during the last 100 years. It is an admitted fact that Continental railways are not as good and safe as English railways. In France there are more accidents on the railways than there are in England. In France an engine driver is paid according to time he takes to make a run, and receives a bonus if he makes the run above the scheduled time, of course this encourages him to speed and consequently heighten the risk of accidents.

The speed of the French taxi is famous, but surely this is not a good reputation, it is much better to travel a little slower and so lessen the risk of accidents. Although the death rate from road accidents is high in this country I think it compares favourably with other countries.

Concerning the question of Empire, it is surely a fact, that we possess the greatest Empire the world has ever known, and the bed rock foundation of it in my opinion is the king and monarchy—coupled with democracy. Our Empire will last forever in my opinion, if it remains on its present basis—it is the greatest factor in world peace, and while the Dominions of Canada, South Africa, and Australia, have their own governments, and are allowed to manage their own affairs, they will cling and support the Mother Country for ever.

\* \* \* \* \*

Then, as number of things in this country in this country want "nationalising" and they want it badly,—coal mines, gas, electricity, water, and all communications want to be brought under government control as they are in most continental countries. Our farming is in a mire, it wants fertilising,—and nationalising as it is in Denmark,—we could become much more self supporting and independent if we made a "Return to the Land" as France, Italy, and Germany are doing.

Even so, England is considered by most people who have travelled and explored to be "the best country in the world." Our democracy, liberty, of the press, speech and religion, are a national heritage and knock dictatorships and despotism into "cocked hats." We must safeguard and uphold these rights, and never let England down, but keep it the best, most free and safest country in the world.

No. 58.

ESSAY B1  
IN PRAISE OF ENGLAND

Any English man who has been abroad for some time will find many things to say in praise of England. Almost any man would have this feeling for his Native country, but with the English people the feeling is supposed to be greater. Although people go abroad for holidays, they are always glad when they see the familiar English countryside, and hear the country accent.

England started very insignificantly, a very small island "Britannia," it was almost wholly ignored by the Romans, and the first real conqueror was Hadrian. Then later, after the Romans had left, Scots and Danes invaded it, plundering and destroying. William the Conqueror brought in French blood, and England was really a mixture of warlike people. And so England grew from an isle inhabited by savages, to a land with a huge Empire, one of the foremost in the world.

English soldiers have won fame everywhere and, as Napoleon said, "They don't know when they are beaten." It is this that has made England's name for her. The bulldog is a good emblem of British pluck, as it holds on and won't let go of its object. The British Navy is the oldest in the World, and is called the "Senior Service" as it was formed before the Army. Alfred the Great founded the Navy to guard against the Danes and Vikings. The Royal Air Force is very new, but it had a chance to distinguish itself in the World War. Part of it acts as policemen in Eastern provinces, and has helped a lot in subduing risings and rebellions.

The British Police are a model to any other country, and although the English policeman is depicted as a fat blockhead, he is not so by any means. They help to make English towns what they are, and although "coloured lights" are used for regulating traffic now, England cannot do without policemen. They help to make London the first city in the world.

As a contrast to towns, there comes the English countryside. Many poems are composed about it, and many pictures painted of it. In different parts of England, there are, naturally, different types of country. In Kent there are many hops, orchards etc., while in the Midlands there are many sheep. It is a very beautiful sight to stand on a hill and gaze down on different fields. Some are yellow with corn, some green grass with cows on it, and some with vegetables etc.

The greatest blot on England are the "slum" and "factory" areas. These were started and made half a century ago, but now England is doing her best to clear out these "special" areas. Carlyle attacked the "get rich quick" industrial manufacturers in his essays, and greatly helped to bring about a great change. The employee at a factory today is greatly helped by the "Unions," and the special areas are being changed.

The British Monarchy is one of the oldest in the world, and the king-emperor is the first person in the country. The loyalty of the subjects is never questionable, and the thought of the King helps to bring about many a great desire. Although a kingdom, England is a truly democratic country, and honourable.

151. The following are the marks for Sense and General Impression for four essays on subject C:—

	Examiners	P	Q	R	S	T	U	V	W	Range
No. 31	Sense	12	11	11	8	10	12	10	10	4
	General Impression	11	10	9	8	10	11	10	7	4
No. 60	Sense	9	10	12	11	11	9	8	8	4
	General Impression	9	11	12	11	11	10	8	8	4
No. 77	Sense	6	12	6	8	11	13	11	11	7
	General Impression	7	11	5	8	9	11	10	9	6
No. 85	Sense	9	12	11	9	9	8	11	8	4
	General Impression	9	12	11	9	9	9	11	7	5

152. We quote below the essays of Nos. 60 and 77 and 85.

No. 60.

### ESSAY C

May 6th 1937.

#### A SCHOOL SPEECH-DAY

*Material.*—Reason for speech days—opportunity for parents to see boys and masters, how school is run. Speeches made by members of company, and chief masters. Distribution of prizes for end of school year. After the speeches, they go down to ground and watch the cricket. Teas in the tents. Spoilt when raining.

- Plan*
1. Reason for school speech-days.
  2. Start of a school speech-day.
  3. Speeches by masters.
  4. Opportunity of parents to see boys and masters.
  5. Prizes at end of school year.
  6. Afternoon at the ground.

#### *A School Speech-Day*

School speech-days are usually organised at the end of every school year, and are held in the mornings, on the last day of the term. They enable the parents and boys who are invited to listen to an account of the doings during the school year. The masters give this news which also includes any new reforms introduced into the school. Also, at the end of the morning, by which time the speeches and prize-giving are over, parents and boys can go down to the ground, where the boys play while the parents watch.

When the parents and boys have assembled in the morning, the speeches begin. The masters of the Company who own the school then make speeches and give some idea of the work of the Company. The Headmaster then speaks and sometimes the Head of the school makes a speech in Latin.

After these speeches the prizes are distributed to boys who have won them during the school year, either for their general good work, or on the results of the summer examinations, held a few weeks before. The prize-giving is rather an ordeal for the boys unless they have plenty of self-confidence.

The afternoon down at the ground is very pleasurable if the weather is good. The parents are at liberty to wander round the ground and

watch the various games going on, of which there are often a dozen on speech day. There is also a side composed of fathers who play cricket, and this plays a side from the school. When the games have finished for the tea interval, everyone is provided for, with the aid of a large tent. After tea, the games continue while some of the parents return home.

A speech-day is quite spoilt when there is rain, for the parents have no opportunity of seeing the ground, and the boys playing cricket. However, the morning speechday is always an interesting affair. The parents can inspect the school buildings and the boys' form-rooms. Thirty years ago these speech days were much the same though the parents then had no opportunity of seeing the ground. The speeches have not differed much, especially the last words of the Headmaster—"The Term is ended." The school is complimented on its performances during the last year, and are exhorted to do even better than this during the next school year. Finally the boys are told to enjoy the holidays and come back ready for a good term's work. Then everyone files out and another school year is over.

No. 77.

### ESSAY C

#### A SCHOOL SPEECH DAY

Practically every school has a speech day some time during the year, and ours usually comes late in November. Personally I've always thought it would be better in the summer, because as it is, it invariably 'pours in torrents,' and literally damps everyone's feelings; certainly a large number of speech days are summer affairs, and from reports I have heard of them, very successful too. Not that ours isn't, but why not take more advantage of the longer summer evenings, and even go the length of having it out of doors? It would at least be original.

However, as it is, we meet every November in either the School Hall, or the Town Hall,—that point depends on finances,—and sing, listen to speeches, and clap at intervals. For hours previously, one can imagine countless mothers scrubbing countless ears and necks all over [*name of a town*], and pushing, and sewing numberless girls of all ages, and sizes into a variety of different white frocks. For we wear white, and look like angels, and by the end of the evening, feel quite the opposite. We are not given programs, so that any paper required for playing noughts and crosses has to be brought in with you, but fortunately we have not had to resort to that type of entertainment, at least as far as I know. By far the most interesting speech day that I have been to was the last one. It was in the Town Hall, and 'hearing' was very much better than usual, which was fortunate the speaker was genuinely interesting, and did not have that irritating habit of 'talking down' to her audience, or making feeble jolks at which one was obliged to titter gently.

By far the most important part of the whole official evening is the singing, upon each one of us depends a great deal, and those who sing 'seconds' feel suddenly, and unexpectedly that they can't remember their note, and much humming goes on under their breath. Actually, of course, every thing goes off perfectly, the music mistress breathes again, and we all collapse into our seats with sighs of relief.

Then follows the Headmistresses report which contains much sound advice to parents, who, poor things, are stacked away up in the gallery, mother wondering if her child's stocking has fallen down yet, after having broken one suspender as she left the house, and father looking interestedly at the staff who endeavour to hanuner sense into his infant.

I often wonder if our staff get a few, not unwelcome shocks, on Speech Day. They meet our parents, and ones character is pulled to pieces, while one fidgets idly, waiting for them to finish talking. Completely new ideas must be formed regarding girls after the form mistress has met the mothers, also she must feel a little different towards her girls after having seen them in long white frocks. This applies notably to the older forms; on ordinary days, in an ordinary brief gym-tunic, a girl looks very little older than fifteen, but on Speech Day, hair done properly, frock long and daintily made, a girl suddenly grows up and becomes someone totally different. That is gain to the staff, but Speech Days not only help them but the parents also, one must not think of such a day as only speeches, songs, and prize giving. There is far more to it than that, mother and form mistress meet, co-operate, and begin to understand the part each must play in the growth of a girl.

But I have tarried long in this part of Speech Day, not a fault my reader will condemn I hope, to my idea, this part is more important than the main 'Prizegiving,' but perhaps that is because I have never received a prize! Even so, one wishes that the girls could play a greater part on the Speech Day, other than by singing. The Head girl makes her speech of thanks to the visitor, and we are all there in our white rows, but sometimes feel that we are looked upon as nice little girls singing nice little songs,—if they only knew!

Therefore one comes to the conclusion that there are two parts to every Speech Day, or there should be if there isn't, the official part, and the unofficial, both important, but the latter most definitely so. So year after year we shall look forward to Speech Day, knowing that one will sing, clap, and listen, and finally hear ones' behavior being discussed, and, frequently condemned.

No. 85.

### ESSAY C

#### A SCHOOL SPEECH-DAY

The thought of Speech-day is first brought to the minds of the pupils when the date of the event is announced, probably after prayers, in the main hall by the head mistress or a mistress deputising for her. Sometimes a few minor details are given at once but these usually come at a later date when they have been gone into more fully.

Preparations for the day begin a considerable time before the event is to take place. The music has to be chosen by the chief organiser and the place for the pupils have to be arranged carefully in order to allow enough rooms for the parents who will attend. When all this has been done there must be several rehearsals of the anthems, songs and musical interludes so that everything shall be cut and dried for the day itself. This means often that lessons must be interrupted at times. These interruptions are sometimes welcome, especially if the lesson is somewhat dull and uninteresting. At other times, however, when

a lesson is interesting or the loss of a lesson will mean that the class will become more backward in its syllabus, the interruption is annoying and inconvenient. This may often occur but if it is unavoidable it must be taken in the best possible spirit.

Life at the period is not always a bed of roses. Annoying little incidents may occur to put one's nerves on edge. A song may have to be repeated again and again before it reaches the standard required by the organiser, if it ever does. Sometimes, the whole school seems to all sense of direction and muddles ensue which have to be righted at the waste of valuable time. Often the return of an absentee puts a whole section out of order or a pupil may take it into his or her head to move his or her position in order to be near a friend.

At last, however, the preparations are complete and the day itself arrives. Last minute instructions are given and minor details arranged. The pupils are marshalled into their places and parents conducted to their seats. Then the distinguished visitor, who is to present the prizes and certificates, arrives and the programme is ready to commence. Sometimes a parent arrives at the last moment and room has to be hurriedly made for her. Then, to the accompaniment of the rustling of paper and the scraping of chairs, the ceremony is opened by a burst of song from the school choir. Usually everything proceeds quite smoothly though the musical interludes are not always given as much attention as they might be given by the pupils.

When the speeches begin they are attended to quite intently at first but after a while, unless the speech is very interesting, one's attention is apt to wander and it is sometimes with a feeling of relief that the moment for presenting the prizes. There then follows a period of prolonged clapping as the pupils go forward to receive their awards while their less fortunate or less diligent associates look on enviously or uncaring.

After the distributing of the prizes their are usually a few more speeches, mostly of thanks to the visitor and then once more the Speech-day is over and the pupils and parents disperse to examine the school buildings or return home to tea, after which the pupils will enjoy an evening free from school work before returning to the never ending round of lessons again.

### 153. The following are the marks for Sense and General Impression of four essays on subject C1 :—

	Examiners	P	Q	R	S	T	U	V	W	Range
No. 4	Sense	12	7	12	11	12	9	9	10	5
	General Impression	11	10	10	11	12	9	9	10	3
No. 12	Sense	6	14	12	10	12	11	11	6	8
	General Impression	8	13	12	11	12	11	12	10	5
No. 15	Sense	8	14	11	8	9	12	11	12	6
	General Impression	8	13	11	8	8	12	12	11	5
No. 88	Sense	9	12	8	8	7	13	10	12	6
	General Impression	8	12	6	8	8	12	10	11	6

### 154. We quote below the essays of Nos. 12 and 15.

No. 12.

### ESSAY C1 OUR LAST SPEECH-DAY

Never before have I been present at a speech-day quite like it, never again shall I attend any function which leaves quite the same impression. I shall never forget it—not because of any extraordinarily interesting speeches, overflowing with the brilliance of the speaker's wit; oh no; there were no exceptions to the rule in that respect! I shall never forget this occasion because of the atmosphere that prevailed throughout it. In describing it as a tense atmosphere of suppressed tedium of even annoyances on the part of the school, of disgust on the part of most of the visitors, and of extreme and obvious embarrassment on the part of the headmaster, I should even then be conveying only a most vague and unreal idea of what it was like. It was an atmosphere peculiar to the occasion and one that could only possibly exist on a school-speech day.

Now that I have attempted to "get you into the thing" so to speak, I will give you a brief account of the proceedings from beginning to end.

The Mayor was to give the prizes and perhaps there was some little more excitement than usual among the younger boys—which lasted till about ten minutes after the ceremony had started.

Soon after we had been jumbled and jostled into our usual corners and back-benches in the hall, visitors started trickling in. They were met by the familiar din and clatter, but when the centre of the hall was practically full of them, we were convinced that they were making practically as much noise themselves.

There was the usual touch of colour, supplied by the masters but even this seemed to disappear as, one by one, or, I should say, two by two, they hid themselves away in corners, either to keep up a just audible running commentary on the ceremony, or else, determined not to have their daily routine disturbed, to take their afternoon nap, which they are accustomed to take in the history period.

Only one or two masters were prancing about on the platform, throwing themselves and their most terrifying voices here, there, and everywhere. It was humour at its best to watch some of them trying to instil ideas of law and order into some of the younger juniors.

(Old A was fussing about as usual).

It was soon twenty five minutes past two, the ceremony was to begin at half-past.

And so, true to tradition and all rules of ceremonial etiquette, at twenty-five minutes to three, the headmaster entered the hall, followed first by the Mayor in all his gilded pomp, and then by a straggling collection of decrepit and decaying governors.

There followed the awaited periods of deep tedium. First, the headmaster's report in its proper, uniform, monotonous tone—although I think he surmounts the difficulty magnificently by rushing through it as quickly as he dare; Secondly, the actual distribution, amidst unending applause (this is practically the only time when the juniors can vent their feelings as they wish); and, omitting other superfluous detail, lastly came the Mayor's great speech.

He opened in the generally accepted way by saying how pleased he was to come and give the prizes (after cursing the whole thing for all the previous week); he went on, amidst deafening applause which now developed into uproarious and hilarious laughter, to say that

he never won a prize when he was at school (also, of course, the generally accepted way of continuing these speeches); and he concluded (still keeping strictly to tradition) by mumbling a tanglo of inaudible rubbish about citizenship and wild oats.

I distinctly heard sighs of relief when the whole thing ended!

No. 15.

ESSAY C1

[10 Bakors Road,  
Southwick.]

26.7.37

Dear Mary,

You did not miss very much by being absent on Speech Day, but I suppose you might just as well be bored now instead. I am going to tell you exactly what happened. The event will probably be in the wrong order, but I can't remember everything. In any case, it was exactly the same as any other Speech Day. We never do get much variety, except that sometimes we have a man as the chief prisoner of the day.

To begin with, we all went into the hall at the back, naturally, and watched our beaming parents fill up the front seats, blocking our view entirely. Then the honoured ones arrived and filed on to the platform. We stood up to get a better view, but the parents followed our example and we were lost again. I do think they might consider us just a little. I believe the prayers and hymns came next—I don't know exactly, but they seem to fit in just here. After that we all sat down again. Then the chairman spoke. We could not hear what he said, but it was an awfully good speech. Somebody thanked him, and then the main speaker started. It happened to be a woman this time. May, I really wished you were there during that speech. I could hear every word and it was very interesting. Above all, the speaker actually told us two new jokes that were awfully funny. All the girls were absolutely staggered and nobody yawned or spoke all the time the speaker was standing.

I can't remember very much after that speech except the prize-giving. There were so many prizes that I seemed to be about the only one not getting one. My conscience tried to give a twinge, but as I haven't slacked any more than usual, it didn't quite work. My brain cannot be as big as other girls', that's all.

After that there were several other speeches thanking the first speakers and so on, followed by the Headmistress's report. Several girls have done very well and got scholarships and all that this year. We were all told to work hard, and then we might possibly do as well as they had done.

Next came the most important thing of all—our play. So that you shan't be prejudiced, I will tell you a few criticisms I was not meant to hear.

"It was a bee-beautiful play, my dear. These little children ca-a-an act well."

"Beastly thing—plays get worse every year, and as for Speech Day ones . . ."

"Very well done, very well done indeed. My daughter was in it. Takes after me. Did you recognise her?"

"Haven't the faintest idea what it was like—slept through it all."

Now to continue. It was something from Shakespeare, and our form did it. Everybody said to us that it was magnificently done and that we were all budding Greta Garbos'. You had better take that with a pinch of salt. The costumes were rather good, and everyone knew her part. I am much too modest to say any more, but you ought to know the form well enough to fill in any omissions.

When the play was finished, we all collected our parents and went to the dining room for tea. Everybody naturally ate piles of food and those serving tea had a beastly job. Parents are in an awfully good mood on Speech Day. Daddy gave me a shilling. I told him he owed me one and he didn't ask what for. I only said it for a joke, but when he actually presented me with the shilling, I certainly was not going to enlighten him.<sup>1</sup>

Do you know, May, I think that Speech Day and Christmas Day are the only days of the year I get any praise on? On Speech Day, the cleverness of the rest of the form reflects on everybody and I received praise for the number of prizes won by the form and for the play. I think that's rather nice, don't you?

On the day after, the whole school was praised for its admirable behaviour and 'ladylikeness'! That was because you weren't there I expect. I hadn't better write any more now, as it might tire you. I think that's a jolly good excuse and way of putting a stop to the letter. I do hope you will soon be better and back at school again. Again I say—I'm frightfully sorry you missed those jokes. I'll tell you them when I see you. Jokes always do look mad and not a scrap funny on paper.

Goodbye for now,  
Yours affectionately,  
[CAROLINE]

P.S.—The latest craze is to pretend to be bored stiff. As a matter of fact, Speech Day was fun, and I think most of us rather enjoyed it, particularly because we could clap so much for no reason at all.

[<sup>1</sup> The Committee are not responsible for the morals of their examinees!]

## CHAPTER VII.—CONCLUSIONS<sup>1</sup>

154. The purpose of the present investigation is plainly stated in the first two paragraphs of the Report: it was to ascertain whether by the introduction of the element of "Sense" with a defined connotation into the marking of essays both the validity and the consistency of examinations on English essays could be improved. It must be stated equally plainly that, while the results of the investigation are in the view of the Sub-Committee interesting and valuable, that main aim has not been achieved, for reasons which we shall suggest. The statistical report states that, although the examiners themselves agreed in regarding the category of Sense as of great value (para. 35) these examiners do not show greater agreement when marking for a well-defined category, such as Sense, than when they mark for General Impression.

155. A first method, which led the statisticians to the above conclusion, was the use of "standard scores" by which we can eliminate the differences of "standard" or "leniency" between examiners as measured by the mean (or average) marks, on the one hand, and the differences of the "spread," or "standard deviation" of the marks from that mean, on the other. The actual (and laborious) method of comparison was made by the drawing of 1200 "cumulative frequency curves" and an examination of their characteristics (see paras. 51–58 above).

156. These results were confirmed by a second method called the "Analysis of Variance," explained in the Note on this subject, on p. 149. In this analysis the statistician attempts to distinguish three components in each mark: the "absolute merit" of the script, the examiner's idiosyncrasy, and his random errors. Dr. Ebblewhite Smith comes to the interesting conclusion that while the absolute merit has the greatest effect in determining the mark awarded, the combined effect of the random error and the idiosyncrasy of the examiner is probably greater than that of the

<sup>1</sup> Reference should be made in this connexion to the valuable passages dealing with English and English Literature, as well as to the general chapter on the School Certificate Examination, in the important report of the Spens Committee (Report of the Consultative Committee on Secondary Education with special reference to Grammar Schools and Technical High Schools (H.M. Stationery Office, 1938, 3s. 6d.), pp. 218–228 and 254–267).

absolute merit for all categories except Sense and General Impression ; and that though these categories were those which were marked " most satisfactorily " the effect of the absolute merit was [only] very slightly larger than the combined effects of the random error and the idiosyncrasy of the examiners—a not very reassuring statement ; and, finally, that neither the examiners' differences nor the random errors of marking are significantly less for directed than for undirected essays. Since it is obviously easier for examiners to mark for " Sense " in the directed essays, in which object and audience are explicitly defined, than in undirected essays, in which they have to be guessed by the examiner, it is again evident that the introduction of the element " Sense " has not had the desired effect on the marking.<sup>1</sup>

157. Dr. Smith, in the sections dealing with Factor Analysis (paras. 86-95), also investigated the marks for the various categories by the use of correlation-coefficients with a view to determining, if possible, the influence both of the various categories, and of the capacities tested by the objective tests, on the marks for General Impression. But he finds the " factor-pattern " difficult to explain. One factor only is sufficient to account almost entirely for the inter-correlation of all the seven categories. On removing this factor, a single group-factor was sufficient to account for the correlations between the objective tests (see para. 92). He came to the conclusion that the high value in all the categories was due to what is known as a " halo " effect, i.e., examiners when awarding a mark for other categories are unable to free their minds entirely from the General Impression, and their marks are really very strongly weighted with General Impression.<sup>2</sup>

157A. Our own general interpretation of the failure to improve the consistency of the marking is that the examiners, brought up in the old traditions of teaching and examining, in spite of their manifest ability and their desire to conform to a new method, were unable to free themselves from those acquired traditions. A careful investigation on a smaller scale by Dr. B. M. D. Cast has led to similar results. (See Note on p. 157 below.)

<sup>1</sup> We have not quoted *in extenso* the conclusions, for which reference must be made to para. 85 above.

<sup>2</sup> Attention should be drawn here to the conclusion of the Note on pp. 150-1, in which it is stated that Dr. Burt and Dr. Godfrey Thomson, to whom the proofs of this book were submitted, have independently suggested that a psychologically more acceptable analysis would be one which produced two factors among the categories, namely a general ability factor and a " halo " factor ; and Dr. Thomson has suggested a method by which this result might be achieved. Dr. Spearman has also made another important suggestion with regard to this matter.

158. Our statisticians in their first analysis have used the term "absolute merit" and have implied that this is capable of being measured by numerical marks. It is a term that needs closer scrutiny. Dr. Rhodes, in *The Marks of Examiners*, has worked out a method for calculating out from the actual marks the "ideal mark" for a "piece of work" by the "perfect" examiner. It is based on the assumption that the average verdict of a number of examiners is better than any of the single verdicts. We may here recall the postulate by Professor F. Y. Edgeworth in his pioneering work on the statistics of examinations that the true or standard mark for any piece of work is the average of the marks given by a large number of competent examiners equally proficient in the subject and instructed as to the character and purpose of the examination. Edgeworth clearly stated that he regarded this as a postulate and not an axiom, and expressly noted that his subject was not the philosophy of examinations. (See the *Journal of the Statistical Society*, vol. li, pp. 599-635 (1888) and vol. liii, pp. 460-475 and 644-663 (1890)). As one of us pointed out in a previous publication (*Examinations and their Relation to Culture and Efficiency*) (1918, p. 103), the weakness of the postulate for practical use is the absence from most examinations of any clear definition of purpose.

159. Dr. Rhodes (*op. cit.* para. 369) remarks (in the language, perhaps, of understatement) that one examiner's idea of "perfection" may be different from that of another examiner, and he attempted to calculate "ideal marks" from a mathematical treatment of the marks actually allotted. Underlying the statements both of Edgeworth and of Rhodes there are fundamental difficulties which we must now approach.

160. It will be convenient to approach them by considering the problem of making some very simple test in a subject less complex than that of English composition, one in which all examiners, without instructions from any examining body, but using universally adopted conventions, must necessarily agree as to the "perfect answer." The simplest possible example is perhaps a test in simple addition; yet even in such a test we shall find that it is easy to encounter at an early stage possible and legitimate differences of opinion as to how the answers should be marked. For if we test only the addition of simple digits of which the sum amounts to more than 9, we shall have to decide whether any answer containing one right and one wrong digit should receive any marks or none. And as we complicate the sums in arithmetic it is only by a complex and rigid system of conventions, on which different experts and examining bodies may easily

take different views, that we can secure any approximation to consistency in the marking of the scripts.<sup>1</sup>

161. A difficulty of another and serious kind arises, affecting other examinations besides those in arithmetic, if we follow a common practice and introduce a non-arithmetical element, such as "neatness," into the marking scheme. It leads at once to an equation such as the following,

5 marks for neatness = 5 marks for accuracy,

though the maxima prescribed for the two categories may be very different. We do not challenge the desirability of giving marks for neatness in marking arithmetic scripts, but we suggest that if the result is not to be misleading they should be kept distinct from first to last from marks for accuracy in arithmetic. There can be no rational method of equating the two kinds of marks and the sum of such heterogeneous quantities has no rational meaning. We shall find the same kind of difficulty, together with fresh ones, in the marking of English compositions, to which we now return.

162. We begin by quoting a passage written some time ago by one of the members of our Sub-Committee. "An essay (said Dr. Ballard) is an intricate mental product which can be analysed in a variety of ways and yet can never be analysed completely. Each examiner consciously or unconsciously (as a rule unconsciously) makes his own analysis, measures each element by his own rod, weighs each factor in accordance with his own scale of values, and finally arrives at a verdict which sums up a number of hazy personal impressions."<sup>2</sup>

To diminish the "haziness" of such categories of impressions it is common for examining bodies to classify them beforehand and to define for the several examiners the maximum marks which they are to allot to each category, with a view to reducing the differences in the marking by different examiners. The results recorded in *The Marks of Examiners* amply demonstrate the insufficiencies of the method; but for the moment we are considering not so much the question of the inconsistencies of the method as of its validity.

163. In the first of our own investigations recorded in *The Marks of Examiners* on essays the system adopted followed exactly or almost exactly the scheme for a Special Place examina-

<sup>1</sup> For examples of inconsistencies in the marking of scripts in simple arithmetic see *An Examination of Examinations* (paras. 30-38) and *The Marks of Examiners* (pp. 68-116). The designers of "objective tests" in arithmetic have used great ingenuity to surmount the difficulties but are limited in the type of questions that they can set.

<sup>2</sup> *The New Examiner*, by P. B. Ballard (Hodder & Stoughton, 5th impression, 1929), p. 62.

tion prescribed by the authority which furnished us with the scripts. The categories, with the maximum allotted to each and the (hypothetical) marks of two candidates who would have been treated as of equal "merit" at their scholarship examination, are set out in the scheme below.<sup>1</sup>

## MARKS

<i>Categories</i> <sup>2</sup>	<i>Maximum</i>	<i>Awarded to Script A</i>	<i>Awarded to Script B</i>
Vocabulary	7	1	5
Accuracy	7	3	3
Craftsmanship	7	5	1
Consistency	7	7	7
Completeness	7	1	5
Substance	7	3	3
Quality	7	5	1
		—	—
		25	25

Without further analysis of detail we may ask : Is it possible to attach any rational meaning to the conclusion that two such scripts should be regarded as equal in "merit"? It seems doubtful.

164. But let us now take a scheme used in a second investigation on another set of Special Place examination scripts, with categories more clearly distinguished, which in our view was more satisfactory in its construction than the earlier one,<sup>3</sup> and consider again the comparison of two scripts to which equal marks, amounting to 60%, might have been awarded in different ways.

## MARKS

<i>Categories</i>	<i>Maximum</i>	<i>Awarded to Script A</i>	<i>Awarded to Script B</i>
Quantity, quality and control of ideas	50	23	38
Vocabulary	15	12	7
Grammar and Punctuation	15	10	4
Structure of sentences	10	5	6
Spelling	5	5	3
Handwriting	5	5	2
	100	60	60

<sup>1</sup> See *The Marks of Examiners*, paras. 172-193.

<sup>2</sup> It should be said that the meanings to be attached to the categories are more closely indicated in the Instructions for Examiners given in the book quoted, p. 112.

<sup>3</sup> See *The Marks of Examiners*, p. 118.

Candidate B is clearly regarded as superior in ideas, A in what is sometimes called the "mechanics" of composition. But what does the equation of the marks mean? To equate 5 marks for "quantity, quality and control of ideas" to 2 for spelling + 3 for handwriting, though it is more satisfactory than the kind of equation quoted above, is only a sacrifice to a customary convention in this kind of examination.

165. We suggest that we should eliminate (for entirely separate marking, if thought desirable) those elements of spelling and handwriting which would disappear if the composition were delivered, or read aloud, and of which no account should be taken in judging its intellectual or artistic value. To equate marks for entirely heterogeneous elements like spelling or handwriting with marks for elements like "Sense" or even Sentence Structure is, in our view, an absurdity.<sup>1</sup> Revolutionary as it may seem, we suggest that marks for spelling and legibility of handwriting should at no stage be added to or subtracted from marks for the other elements of a written composition. If the marks are high, they should not be allowed to compensate for deficiencies in other elements of the composition; if they are low they should not be allowed to detract from the valuation of its intellectual qualities. They may be of great importance; if that is so, it should not be beneath the dignity of an examining body to state them definitely on its certificate for the benefit of a prospective employer to whom they may be of use.

166. We come to the more delicate and difficult subject of punctuation, that rough translation to the eye of the pauses, and indeed also the variations of pitch, intensity, and tone, those natural and mainly unconscious devices by which the speaker marshals his utterances so as to convey unmistakably his thought and his feelings to the listener. It represents to the eye one of the most essential elements for the comprehension of organised speech. We may hope to persuade examiners to eliminate from their marking for General Impression the consideration of mistakes in spelling and inelegances of calligraphy. It would, we believe, be impossible to eliminate from their judgments the delays and uncertainties of understanding produced by defects in punctuation.

<sup>1</sup> Spelling can be marked far more effectively by objective methods than in a composition. Many of the mistakes made in writing a composition at high speed in an examination-room may be due to mere inadvertence.

The inspection of a very large number of the thousands of essays submitted to us revealed the interesting fact that nearly 100% of them were perfectly legible, and legibility is the sole point to be considered in judging the handwriting of an essay. Any classification according to "merit" of pupils so nearly equal would be sheer waste of time.

We believe, nevertheless, that objective tests in spelling such as the one we have used, will yield more consistent results than any incidental marks awarded in the judgment of a composition.<sup>1</sup>

167. We return to the actual categories which were settled at the beginning of our work, after discussion by the experienced examiners who took part in the investigation, as those that they wished to take into consideration in the marking, viz. (I) Sense, (II) Spelling, (III) Punctuation, including Formal Paragraphing, (IV) Grammar, (V) Vocabulary, (VI) Sentence Structure and —to be marked *last*—(VII) General Impression (see para. 28 above). We exclude here the category of Literacy, which is not of a parity with the others and which we discuss later (see para. 180 below).

We deal first with categories (II) to (VI). Tables (II) (1) to (II) (6), (pp. 24–29) and the Tables for the Re-marking, XVII (1) to XVII (6) (pp. 61–66) show the considerable discrepancies that occurred in the marking for these categories. The two categories regarded as of the greatest importance were "Sense" and "General Impression," and these were marked, with intention, first and last: Sense first, because we wished to avoid concentration on details in judging whether a composition fulfilled its object or not, whether it "succeeded" or "failed"<sup>2</sup>; General Impression last, because we wished the examiners to bear in mind, though without using the process of addition, all the factors which had contributed to that General Impression (see para. 35). But we purposely refrained from asking the examiners to give any long consideration to categories II to VI which would have involved an expenditure of time out of proportion to its utility.

We defined at the beginning of this Report (paras. 1–6) the special connotation which we have attributed to the word Sense in the marking of compositions, and we believe that when the principle of demanding Sense in the marking of a composition has been more widely adopted in the class-room, it will be more easy to adopt it in the examination-room.

It may be noted here incidentally that the examiners definitely renounced the idea of giving as a final mark for General Impression the sum of marks given to the other categories and thus

<sup>1</sup> We are assured on good authority that, some years ago, specialist examiners at a certain university proposed to disqualify a candidate for the gold medal in a branch of a certain examination because he began many of his lines with commas. Their colleagues assured them that many gold medallists had committed far worse blunders in English, and the specialists relented.

<sup>2</sup> We quoted at the very outset of this Report, as an extreme case, an accumulation of faultless sentences of which the resultant was nonsense. We may have an accumulation of faulty sentences of which the resultant is excellent sense.

avoided the equations of heterogeneous quantities criticised above.

168. The examiners finally recommended the modified set of categories quoted in para. 35A above, viz. (i) Sense, (ii) Spelling, (iii) Punctuation (including Formal Paragraphing), (iv) Grammar and Syntax, (v) Accuracy of Vocabulary (the examiner bearing in mind only whether the words are used correctly or incorrectly), (vi) Power of Expression, to cover (a) felicities of vocabulary and (b) Sentence Structure, (vii) General Impression. This new list of categories, based on the experience gained in the investigation, seems likely to be an improvement on the previous one, especially in three respects : the union of Grammar and Syntax ; the introduction of the term Accuracy of Vocabulary, restricting marking under this heading solely to the question whether words are used correctly or incorrectly ; and the introduction of the term Power of Expression, etc.

On two points to which we now turn we differ from the proposals of the examiners.

169. (i) The examiners recommended an 8-point scale for the marking of categories (iii), (iv) and (v). We do not object to the scale being less extended than for the other categories. But there are decided advantages in a scale with a middle or "median" point and we therefore should prefer a 9-point scale if the 15-point scale is abandoned. It is convenient to make the median point correspond either to an average mark, or, as we have done, to a competency mark.<sup>1</sup>

(ii) Although, owing to the failure of most English schools to insist on the use of a plan, and the failure of examiners to demand it, the attempt to persuade our examiners to mark for Plan failed (see paras. 30 and 34 above), we attach great importance to its inclusion among the categories. We regard the general omission of the Plan from the teaching of the mother-tongue in most of our schools as one of the most serious defects in our educational system. It is true that the novice may find it difficult to think out and plan a composition before he begins to write. But any pupil can be taught to make a summary of what he has written, to criticise his composition in the light of his own summary, and if necessary, to re-write it. A few of our scripts show that this method is already in use by some teachers.

170. The categories actually recommended by the examiners and by ourselves are of two kinds, those to which universally

<sup>1</sup> In marking by impression only it has been found convenient to sort scripts first into three heaps, poor, medium, and good ; and to push the classification further by dividing each of the sets into three fresh heaps in a similar way.

accepted conventions apply, and the rest. The spelling of a composition may be "perfect" and have therefore "absolute merit." That is more or less true of its punctuation and formal paragraphing, its grammar and syntax, the accuracy of its use of vocabulary and its sentence structure. But it is untrue of a plan, still more untrue of "power of expression, covering felicities of vocabulary." What one examiner regards as felicitous, another, with equal justification, may regard as unfortunate. We are launched by that phrase on a sea of uncertainties, the uncertainties of all æsthetic judgments, in which the *terra firma* of "absolute merit" may be undiscoverable. It should be the business of the examining authorities to make sure that no candidate at a School Certificate examination is wrecked on that sea.<sup>1</sup>

171. Now uncertainties appear to be no less conspicuous in the marking of French compositions by French examiners. In investigations carried out by our colleagues in France, the *Commission française pour l'Enquête Carnegie*, three essay-scripts (*copies de composition française*) which had each been awarded 36 marks out of 80 (or 45 per cent.) and had been ranked as 24th out of a batch of 50 at a *baccalauréat* examination were re-marked independently by 76 examiners. The marks for one script varied from 4 to 52, for the second from 12 to 64, for the third from 16 to 56, out of a maximum of 80. The mean marks for the three scripts were 25.9%, 40%, and 34.4%. In another investigation on 100 scripts in French composition presented at the *baccalauréat*, which were marked by 5 examiners (in addition to the one who had marked at the actual examination), 20.1 per cent. of the total number of differences between examiners exceeded 25 per cent. of the maximum marks; and 70 per cent. of the scripts were marked as "passes" by some examiners and failures by others.<sup>2</sup> We suggest that with variations of this kind,

<sup>1</sup> In the investigation on the marking of fifty English essay-scripts presented originally for the entrance scholarship examination of a group of colleges at a university, the most generally lenient of the five examiners placed in the Fourth Class an essay which one colleague placed in the First and the other three in the Second; and there were other striking discrepancies of judgment, especially in the award of First Classes (see *The Marks of Examiners*, pp. 145 et seq.).

<sup>2</sup> See pp. 77-81 and 125-6 of the massive volume published by the French Committee under the editorship of their President, M. Auguste Desclos:—*La Correction des Epreuves Ecrites dans les Examens, Enquête expérimentale sur le Baccalauréat* (Maison du Livre, Paris, 1936, 4to, pp. 387). This volume contains elaborate reports, statistical and general, by distinguished specialists, on the marking of actual scripts furnished at the first and second parts of the *baccalauréat* examination in French composition, Latin translation (unseen), English, mathematics, physics, and philosophy. Details of the *baccalauréat* examination are given in another publication of the French Committee, *Atlas de l'Enseignement en France* (Maison du Livre, Paris, 1933). The two parts correspond roughly to the School Certificate and Higher School Certificate examinations.

and of those shown by our own investigation, the mean marks of a large number of examiners desiderated by Edgeworth have no real significance. The examiners are obviously evaluating different elements in the compositions.

172. But if candidates ought not to be ploughed by chance, neither ought they to be passed by chance. It is obviously the business of the examining authorities to see that a "pass" in English at the School Certificate stage means the possession of a utilisable skill in the use of the mother-tongue, and that it is not liable to be the faree condemned by the Investigators of 1931.

Some of the difficulties in this matter, too complex to be discussed fully in this place, are created for examining bodies in part by their own regulations.<sup>1</sup> But we think it would be reasonable that the candidate who has passed in other subjects, but not in English, should not be required to pass in those other subjects again, but allowed an increased portion of his time to enable him to make good the deficiencies in a skill likely to be of major importance for his future career. Does not the requirement to pass a second time in a subject in which a candidate has passed already cast a doubt on the efficiency of the examination? No one would think it reasonable to require a pupil who had obtained a certificate for swimming 100 yards to undergo the same test a second time merely because he had failed in (say) a horse-riding test on the same occasion.

173. The examination of essays in the form in universal vogue until recently at School Certificate examinations has been criticised by members of our Sub-Committee on various occasions and for various reasons, of which one was its unreliability. "The examiner's blunders," wrote Dr. Ballard, "are as carefully hidden from the public gaze as the doctor's blunders under the tombstones."<sup>2</sup> But one examination authority, the University of Durham, during the early days of our own investigations on examinations had the courage to reveal the glaring inconsistencies of marking in English at the School Certificate stage.<sup>3</sup> Although the Investigators of 1931 were obviously unaware of the inconsistencies of marking to which the University of Durham drew

<sup>1</sup> For a fuller discussion of these difficulties see the chapter on English Composition at the School Certificate Examination in the *Essays on Examinations* published by our Committee (pp. 131-42) and the address by Sir P. Hartog on *Secondary School Examinations . . . with suggestions for Reform* (published by the National Union of Teachers, 1937), especially pp. 17-25, including a summary of the scheme of Dr. W. Edwards, sometime H.M. Inspector of Schools.

<sup>2</sup> See *The New Examiner*, p. 65.

<sup>3</sup> See *The Marks of Examiners*, pp. 64-67; and papers by Mr. C. Roberts and Professor H. V. A. Briscoe, (then) of the University of Durham, published in *A.M.A.* (the organ of the Assistant Masters' Association) for December, 1931, and February and March, 1932, and in the *Journal of Education* for April, 1932.

attention they had detected other grave imperfections in the examinations in English to which the public became alive ; and the Board of Education has recently induced the other seven School Certificate authorities to follow the example that had been set by the Joint Matriculation Board of the Northern Universities, and to make a pass in English (apart from English literature) compulsory at their School Certificate examinations.<sup>1</sup>

Under their new regulations the examining authorities have introduced certain reforms for which the public must be grateful, but of which the precise effect cannot yet be judged. The errors of their examiners may still be concealed under the tombstones of wrong rejections and the monuments of wrong passes.

174. First and foremost of the reforms are the separation of the examination in English language from the examination in English literature and the rendering of a pass in English compulsory, referred to above. Secondly, there has been, in our judgment, a marked all-round improvement in the detailed questions on English in the papers other than those which include the "essay" or its equivalent. Some papers, though not all, include the requirement to write a letter with a definite object in view.<sup>2</sup>

Thirdly, a number of the authorities have abandoned the use in their examination papers of the word "essay" with its tradition, so often fatal to schoolboys and schoolmasters alike (though not to a Goldsmith or an Addison), crystallised in Johnson's definition of the essay as "an irregular, ill-digested piece."<sup>3</sup>

175. It would, however, be a mistake to imagine that the old tradition of asking pupils of 15 to 16 to write compositions on vast or vague subjects (or both) for the world at large has died the death it deserves. Here is a list of some subjects set in 1939 and 1940 at School Certificate examinations :—

Sugar.	The North Sea.
Cathedrals.	Pockets.
Spies.	Poland.
Reading.	Beasts of Burden.
Wheels.	Sunday.
Wanderers.	Money.
Etiquette.	Marvels of the Sky.

<sup>1</sup> See the references to the Report of the Investigators and the action of the Board on p. 80 above ; and also *Essays on Examinations*, pp. 131–142.

<sup>2</sup> The following is an example of a question actually set :—" Write a letter (of about 100 words) to a railway company complaining that your luggage, sent 'in advance,' to the place where you are spending a holiday, has not reached you and is already two days late. Give all necessary particulars." Why should not a letter of this kind (perhaps rather longer) be regularly set as part of the English test ? It would have an excellent influence on the class-room.

<sup>3</sup> Quoted in the Shorter Oxford Dictionary.

Imagine the difficulty of equating the merits of an essay on Poland with those of one on Pockets or Sunday. The tradition of the "ill-digested" pieee stares one in the face. We know what to expect from most schoolboys or schoolgirls asked to write on such topics in the time allowed in the examination-room. It is only fair to say that many of the other subjects are on a different level and seem to us to show a greater degree of common-sense in their selection.

176. The time allowed for the "essay" or its equivalent is not always exactly specified—but in the regulations in which it is actually specified it varies from three-quarters of an hour to an hour and a quarter.

177. One authority is obviously anxious to reduce that time to a minimum. In an essay-paper it gives the following instructions:—

"Remember that regard will be paid chiefly to soundness of style and suitability of material, and not to mere length. A good composition of a page and a half, in handwriting of average size, will be sufficient to earn a high mark, and your composition must not exceed three pages in length."

Obviously the authority regards a serious test of the power of composition and literary construction as not to be thought of in its examination, probably in view of the uncertainties of the results. Nevertheless the terms "soundness of style" and "suitability of material" may seem to need some further elucidation. A style that might be sound and the material that might be suitable with one kind of audience and object in view might be no less unsound and unsuitable for another. It is needless to enlarge on this criticism which raises again the general problem of purpose in writing.

178. Two reforms in the tests of English composition besides those already suggested might be useful at the present day. The Board of Education, through its Secondary Schools Examinations Council, might request the examination authorities to furnish them with specimens (chosen by the Council from the sole inspection of the numbers and marks of the candidates) of English scripts to which the borderline mark (*a*) for a pass and (*b*) for a credit have been assigned; and some of these specimens might be ultimately published for the information of the public.

Further, the Board might ask the authorities to test the consistency of marking of their own examiners by requiring them to re-mark at the end of the examination, say, 5 per cent. of the scripts that they had marked in the first instance, all indications of the numbers and first markings of the scripts having been

removed (as was done in re-marking at the present investigation); and information on the results of these tests should be transmitted to the Board. All concerned would then know more of the accuracy of the methods on which the fate of so many candidates depends. We have reason to believe that tests of this kind have been made by more than one authority. But the results have not been revealed to the public.

179. We venture, not without hesitation, to make one still more drastic proposal. It is that the lengthy composition, so difficult to judge fairly in the examination-room, be *temporarily* dropped altogether from the School Certificate examinations. The objection is obvious, that this fundamentally important means of stimulating both creation and criticism, if properly used, might disappear from the schools, with unfortunate results. There may, however, be a way out, namely, that the Board of Education should lay down conditions in its rules for grants requiring that a sufficient amount of time should be devoted to the subject; and we might trust to teachers and inspectors to see that this time was properly used. The composition would replace the essay; the consistency of tests in English would be increased in the examination-room; and we may hope that the efficiency of the teaching would be increased in the school.

180. As a postscript to the summary given in the preceding pages we have to add a brief abstract of the chapter on Literacy. We pointed out that the report of the Government Investigators of 1931 suggested, though not in so many words, that the term Literacy might be used as a touchstone to distinguish between candidates who should be passed and those who should be ploughed at the School Certificate examination.<sup>1</sup>

The suggestion was attractive and we hoped that it might prove fruitful. But it has failed us. The variations of the marks were decisive. They show that at the School Certificate stage Literacy has no meaning sufficiently clear and well-recognised in the minds of examiners to be used effectively in judging the scripts of candidates. At the first marking of the 600 scripts there were great divergencies between the different examiners. The results of the re-marking repeated and emphasised the uncertainties of the first; they showed that the examiners disagreed not only with each other but with themselves.

<sup>1</sup> See p. 80 above. No further report of the same kind on School Certificate examinations has been issued up to the present, although a report on the Higher School Certificate examination was issued in 1937 (see footnote on p. 154 below).

In the re-marking of the 120 scripts they changed their own verdicts in 159 cases altogether out of 960, or one in six.

Though our hopes were not realised the results of the tests had a certain interest. They show the degree of attainment reached by certain pupils whom the great majority of examiners regarded as illiterate. On p. 88 we have reprinted *in extenso* an essay by candidate No. 43, the only essay out of the 600 regarded by all the eight examiners as illiterate. It shows that the author has ideas, and can use words so as to convey them to other people. He is the master of a fair vocabulary and his spelling mistakes are neither numerous nor really misleading. His punctuation (including formal paragraphing), grammar, and sentence structure leave much to be desired. But he knows what he means ; and with good teaching he should be capable within a reasonable time of writing correct English and of expressing himself with some force. He is unlikely to become (we quote his own amusing phrase) a "cosin" (he means cousin) to a clam.

In what precedes, we have pleaded that a candidate who has failed in English but passed in other subjects (for which a knowledge of new "periods" or new "special books" might be required), were he to take the examination again, should be excused those subjects so as to enable him to devote more time to the mother-tongue. The case of a candidate with a standard in English comparable to that of No. 43 illustrates our point.

## CHAPTER VIII.—SUMMARY OF RECOMMENDATIONS

181. With a view to the improvement of the teaching of the mother-tongue in our schools and the examinations in this subject which at present so vitally affect the teaching, we make the following recommendations :—

(1) That the practice of asking pupils from the age of 13 and upwards to write “compositions” termed “essays” be abandoned ; and that they be asked instead to write compositions on subjects about which they may reasonably be expected to have a fund of ideas and a sufficient knowledge which they could express for a given audience and with a given object in view ; and that before the writing is actually begun the audience and the object be either defined by the teacher or the examiner concerned, or left to the pupil himself to determine. (See Chapter I and *passim*.)

(2) That in order to teach the pupils how to present their thoughts in orderly sequence they be required when writing a composition of any considerable length to submit with it a short plan or summary. (See para. 169.)

(3) That teachers and examiners be advised to judge a composition in the first instance for its “sense,” basing their judgment on their estimate of the degree of success of the writer in the attainment of his object, and bearing in mind that while an accumulation of perfect sentences may make nonsense, an accumulation of sentences by no means impeccable may make excellent sense. (See Chapter I and *passim*.)

(4) That compositions of pupils of, say, from 13–16, be marked for the following categories, in the order indicated : (I) Sense ; (II) Spelling ; (III) Punctuation, including Formal Paragraphing ; (IV) Grammar and Syntax ; (V) Accuracy of Vocabulary (the only point taken into consideration being the correct or incorrect use of words) ; (VI) Power of Expression, to cover (*a*) felicities in the use of vocabulary and (*b*) sentence structure ; and (VII) General Impression ;

And that the mark for General Impression be allotted after the marks for the other categories, but not by any process of arithmetical summation of those marks. (See paras. 161-8.)

(5) That marks for such categories as spelling and legibility of handwriting be neither added to nor subtracted from marks for any other category ; but that, when importance is attached to them, such marks be recorded separately on any certificate given to the pupil or issued to the public. (See paras. 161-5.)

(6) That provision be made in the curriculum of schools for exercises to be set, say once a fortnight, for the writing of a composition of a length which will require not less than 2 (or  $2\frac{1}{2}$ ) hours' work (including time for corrections and re-copying, if necessary) ; and that notice of the subjects of such compositions be given to the pupils, as a general rule, a week or ten days beforehand.

(7) That the authorities for School Certificate examinations be invited (a) to test the self-consistency of their examiners by requesting them at the end of an examination to re-mark, say, 25 or 50 of the scripts which they had marked originally, all marks on those scripts (including the original numbers) having been removed ; or, as an alternative, that they be asked to mark typed copies of the originals (see paras. 97-105) ; (b) to test the general consistency of the marking of the different examiners by assigning to them at the end of an examination for re-marking a set of 25 or 50 typed copies of the original scripts ; and that one or more sets of scripts be assembled for this purpose, each set being composed of scripts which had been originally assigned to a number of different examiners. (See para. 178.<sup>1</sup>)

(8) That in order to test the real significance of the awards, such as Pass, Credit, etc., for English Composition given by School Certificate Examination authorities each such authority be requested to submit, at intervals to be determined, for the information of the Board of Education, through the Secondary Schools Examination Council six specimens of compositions to which the borderline mark for each of such awards has been finally assigned ; and that such scripts be selected by the Secondary Schools Examination Council by the inspection not of the scripts but of a numbered list of the candidates to whom borderline marks have been allotted. (See para. 178.)

(9) That, for a period to be determined, the demand to write a composition exceeding (say) 200 words in length be omitted

<sup>1</sup> This Recommendation is mainly based on the general statistics recorded in this book as a whole ; see also *The Marks of Examiners*, paras. 662 and 669.

from School Certificate examinations ; and that this restriction should not involve a corresponding restriction in the teaching of English composition, which should always have ample time definitely allotted to it in the school time-table, distinct from that allotted to other branches of English, such as grammar and literature. (See para. 179.)

(10) That in School Certificate examinations the use of précis be extended to passages of not less than 1000 words ; and that the candidates be required to make a précis of such passages not exceeding one-sixth of the length of the original ; and further that the object of the précis be defined as a statement of what the candidate regards as the main points in the original passage for the benefit of a person who has not the time or opportunity to read that passage for himself.

(11) That at the School Certificate examinations no candidate who has failed in English, but has passed (or obtained credit) in other subjects, be required to re-present himself for re-examination in such other subjects, and that if a candidate has passed (or obtained credit) in English he be not required to re-present himself in that subject. (See para. 172.)

## NOTE ON THE ALLOTMENT OF THE COMPETENCY MARK IN THE MARKING FOR GENERAL IMPRESSION, AND ON THE RANGES OF THE MARKS IN THAT CATEGORY

1. In the Note on Certain Points in the Marking of Examiners (pp. 152-155 below) we draw attention to the fact that the "fixed point" from which the marking starts is 50 per cent. in some important examinations designed to test utilisable skills.

2. In the investigation on College Entrance Scholarship Essays, recorded in *The Marks of Examiners* (pp. 142-147), in which 50 scripts were marked independently by five examiners, the important or fixed point for consideration was the limit for a First Class for which the prescribed mark was 67 out of a maximum of 100. But an examiner was permitted, if he so preferred, to mark first for classes and then to arrange his marks to fit his classification; while if he preferred to mark first numerically he could arrange his classification to fit his marks. The discrepancies between the different examiners were great, both in the number of awards of First Classes, and in the selection of candidates for such awards. Though the average marks for the different examiners only varied from 50·6 to 54·8, the number of First Classes allotted by them varied from 2 to 8. Not a single candidate was placed in the First Class by more than three of the five examiners, and only three candidates received three votes each for a First; four received two votes each; and ten had one vote each.

The range (i.e. the difference between the highest and lowest marks awarded to any candidate) varied from 7 to 35, with an average of 20, for the whole set of scripts.

3. In the present investigation the important dividing line for consideration is not a "First Class mark" but the "competency" mark, (8) with a fifteen-point scale. If we could have found any general agreement in the allotment of the competency mark it would have been of real use: we should have taken a definite step in advance towards a satisfactory method of deciding which candidates should pass, and which should fail in English composition at a School Certificate examination. But Tables XXIX

and XXX below, which give the percentages of marks of eight, and of marks above eight and below eight, for each examiner and each essay-subject, both at the first marking of the 600 essays, and at the re-marking of 120 of them, show that the differences between the different examiners at both markings, and the difference between their individual judgments on the two occasions, were great.

4. The percentages of the limiting mark for competency for the five examiners Q, P, R, S, and V vary only from 19.1 to 21.8, while the percentages for the other three examiners, T, U, and W, vary from 11 to 17.5. How far this is due to mere chance is difficult to determine; but it is clear from the inspection of the differences of their figures for the different essay-subjects that there is no kind of constant agreement as to which particular scripts are on the borderline of competency. The differences are made further apparent when we classify the examiners according to their percentages of 8 marks and over, which are as follows:—

P	Q	R	S	T	U	V	W
49.4	78.4	57.6	61.4	45.6	77.0	62.0	62.7

These fall naturally into three groups (P and T), (R, S, V, and W) and (Q and U), quite different from the groups who allot approximately the same percentage of borderline marks.

The following table summarises the percentages of marks of 8, and of marks of 8 and over, allotted at the two markings:—

**COMPARISON OF PERCENTAGES OF COMPETENCY  
MARKS FOR FIRST AND SECOND MARKINGS**

*Percentages of the mark for Competency (8)*

Examiner	P	Q	R	S	T	U	V	W
First Marking	... 19.7	19.1	20.8	21.2	11.0	14.0	21.8	17.5
Second Marking	... 23.3	30.0	20.0	36.7	15.8	22.5	15.8	16.7

*Percentages of Marks of 8 and over*

Examiner	P	Q	R	S	T	U	V	W
First Marking	... 49.4	78.4	57.6	61.4	45.7	77.0	62.0	62.7
Second Marking	... 51.6	82.5	42.5	75.0	84.1	79.2	81.6	62.5

## MARKING FOR COMPETENCY

147

TABLE XXIX  
PERCENTAGES OF SCRIPTS WHICH RECEIVED THE MARK OF COMPETENCY (8), OF A HIGHER MARK (8+) AND OF A LOWER MARK (8-) IN THE FIRST MARKING FOR GENERAL IMPRESSION. (600 SCRIPTS)

Line	P	Q	R	S	T	U	V	W
	No. of 8's 8+'s 8-'s							
17	26	67	21	49	30	18	38	44
16	27	57	20	47	33	22	35	43
12	28	60	19	60	21	19	44	37
18	31	61	23	61	16	17	36	47
22	31	47	16	69	16	24	31	45
33	36	32	17	70	13	25	37	38
Total	... 118	178 304	116	356	129	125	221	254
Average... Percentage of 8 marks and over,	19.720.750.6	19.159.321.5	20.830.842.3	21.240.238.7	11	34.754.3	14	63 23
	49.4	78.4	57.6	01.4	45.7	77.0	62.0	62.7

TABLE XXX

PERCENTAGES OF SCRIPTS WHICH RECEIVED THE MARK OF COMPETENCY (8), OR A HIGHER MARK (8+) AND OF A LOWER MARK (8-) IN THE RE-MARKING FOR GENERAL IMPRESSION. (120 SCRIPTS)

Examiner	P	Q	R	S	T	U	V	W
	No. of 8's 8+'s 8-'s							
5	6 0	7 8	5	4 5	11	6 8	6 0	3 10
4	6 10	6 7	7	2 5	13	6 1	13	4 12
4	4 12	8 8	4	3 5	12	8 8	4	4 14
4	7 0	3 15	2	3 5	12	7 10	3 15	2
7	5 8	9 8	3	0 1	10	8 10	2	2 16
4	6 10	3 17	0	3 6	11	9 0	2	3 15
<b>Totals</b>	28 34	58 36	63 21	24 27	69	44 40	30	19 82
Averages...	23.328.348.3	30.0 52.5 17.5	20.0 22.5 57.5	36.7 38.3 25.0	16.8 08.3 15.8	22.5 56.7 20.8	15.8 65.8 18.3	16.7 45.8 37.5
Percentages of 8 marks and over.	51.0	82.5	42.5	75.0	84.1	79.2	81.6	62.5

## NOTE ON THE ANALYSIS OF VARIANCE AND FACTOR ANALYSIS

[Some additional explanation of the subjects named in the heading seemed to be desirable for the non-statistical reader. The following has been drawn up as a result of consultations with four members of the main Committee, Professors Burt, Hamley, Spearman and Godfrey Thomson. It is regretted that owing to his absence in Canada it was impossible to submit the text of this note to Dr. Ebblewhite Smith for his comments.]

Our results have been analysed by two mathematical methods :—The Analysis of Variance (see paras. 59–85 above) and Factor Analysis (see paras. 86–95 above). The two methods, as applied to investigations of this kind, may be briefly distinguished as follows :—

The Analysis of Variance, in the simple form in which it is here used, is a method of analysing the variation of the marks awarded by a number of examiners to a number of testees on the results of a single test of ability, in order to determine how much of that variation is due (1) to the differences between the testees, (2) to the differences between the examiners, and (3) to random or chance errors.

Factor Analysis as most commonly used may be described as a method of analysing the marks allotted to a number of testees not on a single test, but on a number of tests of different kinds, into different components, so as to show whether two or more of the tests have one or more common factors. The interpretation of the nature of the common factor or factors is left to the person making the analysis and may be made in different ways by different investigators. The mathematical analysis itself tells us nothing about the *nature* of the factors.

For example : A factor analysis of examination-marks awarded to the same candidates in Latin, French, Music and Art, would probably show a common factor, strong in Latin and French, and weaker in Music and Art. This could be interpreted as “abstract intelligence.” It might also show a group-factor involving Latin and French, and another group-factor involving Music and Art. The first of these could be interpreted as a “verbal ability” factor and the second as an “aesthetic ability”

factor. These interpretations are highly subjective, and alternative factor-analyses which explain the data equally well can be made and may be preferred by another experimenter. What we can be sure about are certain relationships which the mathematical analysis shows must in all cases hold between the factors.

The distinction between the analysis of variance and factor-analysis is not, however, as sharp as may be suggested by the above first account. Each of the above situations may indeed be examined by either method, as becomes at once apparent if we think of the examiners of the first situation as being, as it were, so many different "tests" come to life: and "factors" can then be found among the examiners, just as in the second situation among the tests. Moreover, although the analysis of variance and factor-analysis were different in their origins and differ in their mathematical methods, they are not unconnected, and this connection is more in evidence when to the analysis of variance, which analyses the sum of the squares of the deviations in one test, is added the analysis of co-variance, which analyses the sum of the products of the deviations in two tests. It is desirable that both the analysis of variance, and factor analysis, should be used, as here. For the analysis of variance and covariance is more concerned with the fundamental question whether there exist any significant differences at all (as for example between examiners, or between candidates); whereas factor-analysis, the preliminary fundamental question having been answered in the affirmative, seeks to set up a hypothetical structure of "factors" to explain these differences.

Both Dr. Burt and Dr. Thomson have independently suggested, with reference to the analysis in paras. 92-95 of the text, that a psychologically more acceptable analysis would be one which produced *two* factors among the categories, namely a general ability factor and a "halo" factor. Dr. Thomson has suggested that this might be achieved by brigading the categories separately *one by one* with the objective tests, and in a small trial calculation has found encouragement for the view that the general factor loadings in the categories should be lower, and be supplemented by another factor, probably a "halo."

Professor Spearman has suggested a third form of analysis in which the halo is not made responsible for the whole amount of each category but only for about half. It is in fact made to explain only the excess of the category correlations over and above the objective-tests correlations. According to his analysis the categories are analysed into two factors, halo and general, whereas the objective tests show only the general factor. More-

over the general factor admits of further analysis into two well-known and correspondingly illuminative constituents, namely, into the bare general factor (or *g*) together with a verbal factor. In the opinion of Professor Spearman this analysis is as good statistically as the one given in Chapter III and more satisfactory from the psychological point of view.

## NOTE ON CERTAIN POINTS IN THE MARKING OF EXAMINERS

Dr. Ebblewhite Smith, in his Statistical Report (para. 46 (ii)), has drawn attention to the reluctance of examiners to award marks at the bottom and at the top of the marking scale, and the contraction of the scale for practical purposes thereby produced.

The rarity of marks at the bottom of the scale has, we think, a simple explanation. The candidature for an examination involves the payment of a fee ; and hence it is rare to find candidates who know absolutely nothing of the subject in which they present themselves. The knowledge of a candidate may be exiguous, but it is not inappreciable, and examiners recognise this fact.

The reluctance of teachers and examiners in English composition to award the highest mark in their scale, or even one approaching it, is probably due to reasons of a different order. They cling to the old idea that the compositions of their pupils should be compared with some perfect "model," oblivious of the conditions under which that model was produced.<sup>1</sup> If we ask our pupils to imitate not the model but the method of great writers, we may look forward to different results, both in the class-room and the examination-room. We return to our old argument. Let us select for our pupils a subject within their grasp, an audience which they can imagine and realise (somewhat smaller than the whole civilised world), an object in writing that they can attain, and the top of our marking scale will no longer be an inaccessible height. Let it not be thought that the standard implied is a low one. The task of writing is never easy ; and the most experienced of writers can never be assured beforehand of success in any undertaking, though the more competent will often attain it.

We come to questions of another order affecting not only the extremes but the whole of the scale.

The apparent simplicity of numerical marking is, as we have seen before, delusive. It conceals among other things the fact

<sup>1</sup> They are also sometimes oblivious of the obvious fact that the easiest things for a pupil to copy from a model with success are its deficiencies. "If," said a distinguished French critic of an adversary, "If Monsieur X. is not a great writer himself, he resembles many great writers. Like Aristotle, he knows no Latin, and, like Cicero, he knows no German."

that when we are adding marks we are combining something more than numbers and that the sum of our marks may represent something different from our resultant impression.

Here again we are driven back to the question of purpose. In previous publications we have drawn attention to the important distinction between examinations intended to test the attainment of a utilisable skill and examinations intended only to test progress towards that attainment. It is clear that a person who only gets five addition sums right out of ten, and so has earned a total of fifty per cent. of the marks, is no more suited to be entrusted with calculations of the slightest importance than one who gets three or four sums right out of ten. We need not trouble about him if we are looking for a useful calculator. On the other hand, marks of this kind may be of great value in testing his progress and in comparing his progress with that of others. We have pointed out the difficulties which arise in School Certificate examinations when the distinction between these two purposes is overlooked.<sup>1</sup> The examination in English ought to leave no doubt as to what its verdict means. It should declare either that the candidate (1) has a utilisable skill in the written use of his mother-tongue, or (2) has not such a skill, or (3) is considered to be on the borderline.

We have quoted a kind of scale of marking in which fifty per cent. of the marks must be regarded as an undoubted failure. We now turn to another type in which it has been chosen as the "fixed point" for the test of a utilisable skill of the greatest importance to the community, in the degree examinations in medicine and surgery of the Universities of Edinburgh and London. Descriptions of these particular examinations are given in *The Purposes of Examinations, A Symposium*,<sup>2</sup> by Professor T. J. Mackie and Professor F. R. Fraser respectively. Professor Mackie of Edinburgh writes :—

"While the marking is on the percentage basis, this must be broadly interpreted. The examiner has his standard of what constitutes a passable knowledge; to the minimum of this he assesses 50 per cent. and the deviations above or below this level are marked accordingly." (*Op. cit. p. 68.*)

Professor Fraser, of London, is even more explicit :—

"The method usually adopted by the examiners is to give 50 per cent. of the total for each question if the candidate's performance is sufficient and only just sufficient to reach the standard traditional for the degree

<sup>1</sup> *Secondary School Examinations, etc.*, by Sir P. Hartog, published by the National Union of Teachers, Hamilton House, Mabledon Place, 1937, pp. 7-13.

<sup>2</sup> Reprinted from the Year Book of Education (Evans Brothers, 3/-). The book contains a description of 22 important examinations.

and to give more or less marks as the performance varies from this. In judging the answers or performance of a candidate, the attainment of knowledge sufficient to qualify for practice, with something rather more to justify the conferment of the M.B., B.S. degrees of the University, affords a satisfactory 'fixed point' around which the merits of individual candidates can be assessed. With this method of marking, and the high standard demanded for a pass, it is difficult to conceive of any candidate deserving twice as many marks as those required for a pass, and the figures are generally, therefore, on the low side; so that even 75 per cent. is rarely given except in the practicals." (*Op. cit.* p. 72.)

Dr. Fraser goes on to explain that on this system the actual mark first allotted to any candidate considered for the distinction of Honours is usually below that prescribed in the regulations, because of the adoption of a system of marking which was designed for a qualifying examination rather than a competitive one; so that it has become customary to adjust the marks of the candidates selected for Honours so that the regulations of the University may be complied with (*op. cit.* p. 73).

It is interesting to see how a purely numerical system of marking has been modified to suit the actual needs of examinations of the first importance designed to test utilisable skill. It may be added that at the London examination each candidate passes through the hands of fourteen different pairs of examiners.

We shall take as a final example a suggested modification in the allotment of marks at an examination, originally designed as a pass examination, to adapt it at the same time to the needs of a competitive examination. In a recent report on the Higher School Certificate examination the Government Investigators frankly recommended that panels of examiners who thought that a candidate ought to have a State Scholarship should justify that opinion by giving specially high marks to his papers. Commenting on the double purpose of the examination as a pass and as a competitive examination they say:—

"The double function of the Examination cannot be adequately performed if the marking schemes are laid down too rigidly. A marking scheme, adequate and fair for the average candidate, may do rank injustice to the able boy, and a complete catalogue of facts may gain greater credit than an able and original discussion. For this reason it is important for panels of examiners to make up their minds as to the qualities they are looking for in the scholarship candidate and, when they find them, to indicate this clearly either by a system of starring or by the allocation of a high mark, unfettered by too rigid a marking scheme or even ignoring such scheme altogether."<sup>1</sup>

<sup>1</sup> *The Higher School Certificate Examination.* Being the Report of the Panel of Investigators appointed by the Secondary School Examinations Council to enquire into the eight approved Higher School Certificate Examinations held in the summer of 1937 (H.M. Stationery Office, 1s. 6d.), pp. 31-32.

In the three cases just quoted, the allotment of numerical marks is clearly regarded as of use for a preliminary judgment, but serious departures from these marks are made, or recommended to be made, to adapt the examinations to their precise purposes. In examinations on a large scale, far more than in those on a small scale, is watchful inspection of the actual results attained necessary to avoid a contentment based merely on statistical curves.

In paras. 102 and 103 above Dr. Eblewhite Smith has warned the reader that "while correlation coefficients and standard deviations give valuable information on the statistical relation of two series of figures they fail to give some of the most important facts needed by examiners and examining bodies which can only be obtained by comparing the actual marks awarded."

He has supplemented this statement by a paper, written before his departure from England, and published in the *Journal of Education* for May, 1941 (p. 180). He reminds us by a hypothetical example that two examiners might allot marks to six candidates of which the rank-correlation coefficient was 1 (i.e. the correlation was perfect) and yet that, if 33% was regarded as the limit for a pass, 50% for credit and 66% for distinction, they might in no instance award the same mark; that there might be an average difference of 30 marks between them; and that only one candidate might be placed in the same class by both examiners.

He then takes another example of two sets of six marks, again with perfect rank-correlation, but this time with the same general average, and in which the average difference between the marks allotted to the same candidate is 16%, and only two of the candidates would be placed in the same class by both examiners.

Finally, he points out that in an examination in which 20% of the candidates failed, and the correlation-coefficient between two examiners reaches the exceptionally high value of  $\pm 0.95$ , it is to be expected that 7% of the candidates will be "failed" by one examiner and passed by another; and he concludes that in most examinations the element of chance affects far more than 7% of the candidates.

## NOTE ON AN INVESTIGATION ON THE MARKING OF ENGLISH COMPOSITION BY DR. B. M. D. CAST

Reference should be made to an interesting thesis by Dr. B. M. D. Cast, *The Efficiency of Different Methods of Marking English Composition*, of which an abridgement was published in the British Journal of Educational Psychology, Vol. IX (1939), pp. 257-269, and Vol. X (1940), pp. 49-60. The original is deposited in the library of the University of London. Forty essays of the pupils of a central school for girls, aged  $14\frac{1}{2}$  to  $15\frac{1}{2}$ , were marked independently according to four different methods by twelve different examiners. The subject chosen had been suggested by Sir Philip Hartog. It was not, however, a "directed" subject; but at the second marking for "achievement of aim," following the principle suggested by Sir Philip Hartog, the examiners were asked to consider "how far has the writer achieved the purpose which he has implicitly put before himself." Dr. Cast writes (*loc. cit.* Vol. IX, p. 268) "although Hartog's criterion forms an admirable principle for teaching, it is too precarious to form the sole principle for marking." [It is to be noted that the marking is likely to be more precarious when the "object" and "audience" of the writer have to be guessed by the examiner, an uncertainty which is avoided when the subject is directed.]

The research was intended primarily as a preliminary study of methods appropriate to the investigation of such problems. Following an earlier research on the subject by Professor Burt, Miss Cast analysed her data, first of all by the method known as analysis of variance and secondly by the method known as factor analysis: she claimed to show that the two methods supplemented one another and led to much the same conclusion. As regards the conclusions themselves, perhaps the most relevant here is her statement that the most concordant set of marks were those yielded by the analytic method in which the marks were given on a system devised by Burt (*Mental and Scholastic Tests*, p. 331) not fully described in the thesis, and the least concordant by the achievement method (*loc. cit.* p. 261). We would suggest that it is difficult to judge of the results attained in an investigation of this kind without samples of the compositions to which examiners award a high, a low, and a middle mark. Owing to the war it has been impossible to consult the full text of the thesis.

## APPENDIX I.<sup>1</sup>

### OBJECTIVE TESTS.

#### TEST 2. OPPOSITES.

Name of Candidate (*surname in block letters*).....  
 Age..... Years..... Months..... Date of Birth.....  
 School .....,  
 Class .....,  
 Date of Test.....

---

#### DIRECTIONS FOR CANDIDATES

In each question, find among the four numbered words in small letters, the one that is most nearly *opposite* in meaning to the word in capital letters. Put the number of the chosen word in the bracket at the end of the line.

*Example:* FRESH - 1 new 2 insipid 3 stale 4 loathsome - ( 3 )

- |                |   |
|----------------|---|
| 1. SAD -       | 1 serious 2 charming 3 cheerful 4 bad - ( )                       |
| 2. NATIVE -    | 1 coloured 2 relative 3 rustic 4 foreign - ( )                    |
| 3. HOPE -      | 1 faith 2 distrust 3 distress 4 despair - ( )                     |
| 4. CERTAIN -   | 1 doubtful 2 possible 3 careless 4 probable ( )                   |
| 5. COLOSSAL    | 1 disappointing 2 poor 3 small 4 dull - ( )                       |
| 6. ENGAGE      | 1 depart 2 leave 3 dismiss 4 degrade - ( )                        |
| 7. BRILLIANCE  | 1 blackness 2 dullness 3 bluntness 4 squalor( )                   |
| 8. RELAX       | 1 laugh 2 weep 3 tighten 4 bend - ( )                             |
| 9. CLAMOUR     | 1 deceit 2 singing 3 meekness 4 silence - ( )                     |
| 10. KNOWLEDGE  | 1 clumsiness 2 ignorance 3 vulgarity<br>4 stupidity - - - - ( )   |
| 11. BOLD -     | 1 weak 2 desirous 3 timid 4 slavish - ( )                         |
| 12. IMPROVE -  | 1 impair 2 burn 3 invalidate 4 annul - ( )                        |
| 13. ABUNDANT   | 1 small 2 scarce 3 reduced 4 miserly - ( )                        |
| 14. ADMIT      | 1 prevent 2 reproach 3 exhaust 4 exclude ( )                      |
| 15. ARTIFICIAL | 1 ordinary 2 superficial 3 plain 4 natural ( )                    |
| 16. COURTESY   | 1 ignorance 2 thoughtlessness 3 rudeness<br>4 crudity - - - - ( ) |
| 17. GRANT -    | 1 refuse 2 eject 3 exclude 4 grudge - ( )                         |
| 18. IDLE -     | 1 obedient 2 ambitious 3 prudent<br>4 industrious - - - - ( )     |
| 19. FLOURISH   | 1 flounder 2 decay 3 mature 4 age - ( )                           |
| 20. JEOPARDY   | 1 certainty 2 safety 3 escape 4 precaution ( )                    |
| 21. CORPULENT  | 1 thin 2 poor 3 scanty 4 starved - - ( )                          |

<sup>1</sup> Test 1 was a Group-Intelligence test of the National Institute of Industrial Psychology, and is, therefore, not reprinted here. (See para. 24 above.)

22. INDISPUTABLE	-	1 doubtful 2 possible 3 valuable 4 certain	( )
23. KINDLE	-	1 discourage 2 extinguish 3 freeze 4 starve	( )
24. FRAILTY	-	1 rigidity 2 beauty 3 strength 4 brittleness	( )
25. SURPLUS	-	1 arrears 2 minus 3 deficit 4 remainder	( )
26. SERENE	-	1 perturbed 2 displaced 3 disjointed 4 unpleasant	( )
27. MONOTONY	-	1 variety 2 noise 3 excitement 4 delight	( )
28. INJURY	-	1 compensation 2 benefit 3 reward 4 health	( )
29. PROFOUND	-	1 unlettered 2 confused 3 superficial 4 unversed	( )
30. ANCESTOR	-	1 forefather 2 descendant 3 relative 4 progenitor	( )
31. LENIENT	-	1 unkind 2 unsympathetic 3 energetic 4 strict	( )
32. NOBLE	-	1 little 2 base 3 uncivilised 4 poor	( )
33. SUCCEED	-	1 lose 2 recede 3 fail 4 blunder	( )
34. ETERNAL	-	1 infernal 2 seasonal 3 abrupt 4 temporary	( )
35. CONDEMN	-	1 appraise 2 approve 3 retain 4 value	( )

TEST 3. SPELLING<sup>1</sup>

Name of Candidate (*surname in block letters*).....  
 Age.....Years.....Months Date of Birth.....  
 School.....  
 Class .....

Date of Test.....

N.B.—Do not write anything in the column left blank at the right-hand of the page.

1. The room was so low that he could touch the [ceiling] with his hand	Right
2. The royal [carriage] was drawn by four horses	_____
3. There is a considerable body of [evidence] against you	_____
4. With difficulty he [succeeded] in reaching the door	_____
5. The oranges are [delicious]	_____
6. His [secretary] answered most of the letters	_____
7. He was a man who disliked all [foreigners]	_____
8. He had been a [traveller] in many countries	_____
9. She could scarcely control her [disappointment]	_____
10. The doctor gave the patient morphia to [relieve] the pain	_____
11. It was not [sufficiently] well made to stand the strain	_____
12. They showed their [approval] by clapping	_____

<sup>1</sup> A full explanation of the method of administration of the test is given in para. 22 above. In the original form supplied to the pupil the spaces indicated by the dotted lines were left blank for the testee to fill in with the missing words printed here in italics within brackets. The supervisor who administered the test read aloud slowly each sentence in the form completed by the insertion of the word printed here in brackets twice; and the testee was asked then to write it down.

Brought forward

13. I shall [anxiously] await your return  
 14. Your gift was much [appreciated]  
 15. I hope you will [accept] this small gift  
 16. We cannot expect results [immediately]  
 17. He was proud of the new [acquisition] to his collection  
 18. I [especially] asked John to come  
 19. Obtain a [receipt] for all money spent  
 20. She was [recommended] for a scholarship  
 21. Nothing can reverse my [decision]  
 22. I shall have no [opportunity] to see you before Wednesday  
 23. There will be a meeting of the [committee] to-morrow  
 24. She is an [acquaintance] of many years' standing  
 25. It is many years since slavery was [abolished] in America  
 26. In the corner of his garden was an [accumulation] of rubbish  
 27. The window was readily [accessible] by means of a small ladder  
 28. Only a few may [achieve] greatness  
 29. It is easy to [exaggerate] the importance of error  
 30. He fell to the ground [exhausted]  
 31. Such a plan is not [feasible]  
 32. A [ghostly] light shone in the cave  
 33. We have no [guarantee] that you will keep your word  
 34. He had a coloured [handkerchief] in his pocket  
 35. His voice was [hoarse] with emotion  
 36. [Innumerable] rats poured into the castle  
 37. The next [instalment] will appear in our next issue  
 38. It is difficult to work amid [interruptions]  
 39. All men should have [leisure] to read good books  
 40. Few men can tolerate [loneliness]

Total number

## TEST 4. PUNCTUATION AND THE USE OF CAPITALS.

Name of Candidate (surname in block letters).....  
 Age.....Years.....Months Date of Birth.....  
 School .....  
 Class .....  
 Date of Test.....

## DIRECTIONS FOR CANDIDATES.

In each line below, one serious error has been made in punctuation. Where a punctuation mark has been omitted, supply it. Where you find a wrong punctuation mark, or a wrong capital letter, or a small letter instead of a capital, put a ring round it, but do not correct it. It is important to note that only one serious error has been made in each line. Read each passage through as a whole before attempting to make your corrections in the separate lines, as in some cases it is impossible to make the right correction in a line without taking into account the lines which precede or which follow it.

*Example : " Talking of axes , " Said the Duchess,  
" chop off her head !*

You will be told by the Supervisor how to mark this example.  
N.B.—Do not write anything in the column left blank at the right hand  
of the page.

Right

Whenever the horse stopped which  
it did very often ), he , fell off  
in front ; And , whenever it went  
on again ( which it generally , did  
rather suddenly , he fell off  
behind

---

" Take care " cried Alice ; but it  
was too late , the pin , had slipped ,  
and the Queen , had pricked her finger .  
What a thick black cloud that  
is ! " she said , and how fast it  
comes ! Why , I do believe its  
got wings !

---

" If you're going to turn into a  
pig , my dear , " said Alice seriously ,  
" I'll have nothing more to do with you ,  
Mind now !

---

" bless my soul ! " said Mr. Pickwick  
as they stood upon the pavement ; while  
the cants were boing put in . Bless  
my soul , who's to drive I never  
thought of that .

Right

## Brought forward

an old crab said to her son,  
 Why do you walk sideways like that,  
 my son You ought to walk straight." "  
 The young crab replied " Show me how,  
 dear mother and I'll follow your example." "  
 the old crab tried, but tried in vain,  
 and then; saw how foolish she had  
 been to find, fault with her child.

---

A queen bee from hymettus flew  
 up to Olympus with some fresh Honey  
 from the hive as a present to Jupiter  
 who was so pleased, with the gift that he  
 promised to give her anything, she liked  
 to ask for, she said she wanted  
 him to give stings to the bees

---

Demades the orator was once speaking  
 in the Assembly at athens; but the people  
 were very, inattentive to what he was saying;  
 so he stopped and said, "gentlemen,  
 I should like to tell you one of Aesops  
 fables" This made everyone listen intently.  
 then Demades began: " Demeter, a swallow  
 and An eel were once travelling together,  
 and came to a River without a bridge;  
 the swallow flew, over it, and the eel  
 swam across"; and then he stopped.  
 "What happened to Demeter" cried several  
 people in the audience. "Demeter, he replied,  
 is very angry with you. for listening to  
 Fables when you ought to be minding  
 public business.

---

A trumpeter; being captured by the enemy,  
 begged for his life, and said, "do not put  
 me to death; I have killed no-one. Indeed,  
 I have no weapon: but carry with me  
 only my trumpet here.

Total number

## TEST 5. WORD MEANING

[The details of names of candidates, etc., were printed in this place as in previous tests.]

## DIRECTIONS FOR CANDIDATES

In each of the sentences set out below there is a word or phrase which approximately corresponds in meaning to the word in capital letters which precedes the sentence. Underline in each case such word or phrase.

Example : ADORN—Those who came to worship stayed to decorate the shrine with flowers.

N.B.—Do not write anything in the column left blank at the right hand of the page.

		Right
1. ANGUISH	His spiteful rage was appeased at the sight of their distress	
2. DISCUSSION	He excelled in sermon and oration but took no part in quarrel or argument	
3. AWARE	"Careful! My friend, our clever enemy is awake and conscious of our presence"	
4. ADAGE	Age accumulates many a wise saying for the guidance of youth	
5. CELESTIAL	Glorious Spring with sweet voice awakens earth-born birds with heavenly song	
6. DEFRAUDED	He robbed the rich, cheated the poor of their meagre savings, and his untrustworthy evidence in the dock brought him a further conviction for perjury	
7. COGENT	In suitable terms he laid before them logical arguments sufficiently forcible to convince the most sceptical	
8. CONJECTURE	And with this sweep of his arm he brushed aside their doubt; no longer was there need to guess, they were now in possession of the complete argument	
9. FORTITUDE	Arthur displayed his great strength and courage when fighting against great odds in the battle of life	
10. QUERULOUS	The queer old woman was quick and active in her ways, often angry and always complaining	
11. SQUALID	The screaming urchins in the unpleasant streets were a dirty sight	
12. DISPARAGE	Do not despise the cottage or belittle the skyscraper; both have virtues that the wise man may appreciate	
13. ENIGMA	The oracle proposed a riddle in the form of a curious rhyme	

Brought forward

14. EXPOUNDED - While he argued he gesticulated, and as he explained he struck the desk
15. UBIQUITOUS - The wicked are to be found everywhere and even government has its quota
16. FACILITY - He was singularly successful in all forms of sport and he rode and swam with equal ease
17. FEIGN - To pretend to the throne does not mean to imitate royalty, nor to pretend to have royal blood
18. FLEXIBLE - With nimble fingers and incredible skill he fashioned weird figures from the pliable hempen strands
19. INANIMATE - The still silent figure which at first he took to be lifeless proved to be the body of a sleeping man
20. LAPSE - It took him but a few moments to recognise his slip and he fell to his knees and begged his master's forgiveness
21. LOQUACIOUS - He was fond of drinking and in his cups he would become first talkative, then quarrelsome, and finally sleepy
22. MALICE - He was a creature to be despised : he cherished hatred for the police and his treatment of his friends was characterised by spite and ill-temper
23. PATHETIC - The lonely child was miserable and her pitiful weeping roused the sympathy of the passers-by
24. OPTION - You have no choice ; my decision is final and you must accept my conditions
25. TRANQUILLITY - The age of the place and its deserted look gave it an air of expectancy and quietness, so soon to be destroyed by the soldiery
26. TEMERITY - Impudence is often fear and rashness merely foolishness

Total number

## APPENDIX II

### SAMPLE OF MARKING-SHEET FOR ESSAYS.

Sheet No. \_\_\_\_\_

Name of Examiner : \_\_\_\_\_

I.I.E.E. ENGLISH ESSAY INVESTIGATION, CHRISTMAS, 1937

#### MARK SHEET

#### ESSAY (      )

Designation of Candidate.	Sonso.	Plan.	Spelling.	Punctuation, including Formal Paragraphing.	Grammar.	Vocabulary.	Sentence Structure.	General Impression.	Literacy.
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
26									
27									
28									
29									
30									
31									
32									
33									
34									
35									
36									
37									
38									
39									

## APPENDIX II—continued

Designation of Candidate.	Sense. (a)	Plan. (b)	Spelling. (c)	Punctuation, including Formal Paragraphing. (d)	Grammar. (e)	Vocabulary. (f)	Sentence Structure. (g)	General Impression. (h)	Literacy. (i)
40									
41									
42									
43									
44									
45									
46									
47									
48									
49									
50									

Date,.....

Signature of Examiner.....